420 Rec'd PCT/PTO 01 FEB 2000 09/463844

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

REQUEST FOR FILING NATIONAL PHASE OF PCT APPLICATION UNDER 35 U.S.C. 371 AND 37 CFR 1.494 OR 1.495

To: Asst. Commissioner of Patents and Trademarks
Washington, D.C. 20231

(Our Deposit Account No. 03-3975

TRANSMITTAL LETTER TO THE UNITED STATES DESIGNATED/ELECTED OFFICE (DO/EO/US)

Atty Dkt: **PM** 265461 /PHM.70251/UST
M# /Client Ref.

From: Pillsbury Madison & Sutro LLP, IP Group:

Date: February 1, 2000

This is a **REQUEST** for **FILING** a PCT/USA National Phase Application based on:

1. International Application	2. International Filing Date	3. Earliest Priority Date Claimed
PCT/GB98/02259 ✓	28 July 1998 ✓	01 August 1997 ✓
⇧ country code	Day MONTH Year	Day MONTH Year
		(use item 2 if no earlier priority)

4. Measured from the earliest priority date in item 3, this PCT/USA National Phase Application Request is being filed within:

(a) ☐ 20 months from above item 3 date (b) ☒ 30 months from above item 3 date,

(c) Therefore, the due date (unextendable) is February 1, 2000

5. Title of Invention ZGGBP1, NOVEL PEPTIDES RELATED TO BIPOLAR AFFECTIVE DISORDERS TYPE 1, SEQUENCES AND USES THEREOF

6. Inventor(s) FLANNERY, Angela Veronica et al

Applicant herewith submits the following under 35 U.S.C. 371 to effect filing:

7. ☒ Please immediately start national examination procedures (35 U.S.C. 371 (f)).

8. ☒ **A copy of the International Application** as filed (35 U.S.C. 371(c)(2)) is transmitted herewith (file if in English but, if in foreign language, file only if not transmitted to PTO by the International Bureau) including:

a. ☒ Request;
b. ☒ Abstract;
c. 16 pgs. Spec. and Claims;
d. 19 sheet(s) Drawing which are ☐ informal ☒ formal of size ☒ A4 ☐ 11"

9. ☒ **A copy of the International Application has been transmitted by the International Bureau.**

10. **A translation of the International Application** into English (35 U.S.C. 371(c)(2))
a. ☐ is transmitted herewith including: (1) ☐ Request; (2) ☐ Abstract;
(3) _____ pgs. Spec. and Claims;
(4) _____ sheet(s) Drawing which are:
☐ informal ☐ formal of size ☐ A4 ☐ 11"
b. ☐ is not required, as the application was filed in English.
c. ☐ is not herewith, but will be filed when required by the forthcoming PTO Missing Requirements Notice per Rule 494(c) if box 4(a) is X'd or Rule 495(c) if box 4(b) is X'd.
d. ☐ Translation verification attached (not required now).

RE: USA National Filing of PCT/GB98/02259

420 Rec'd PCT/PTO 01 FEB 2000

11. ☒ **PLEASE AMEND** the specification before its first line by inserting as a separate paragraph:
a. ☒ --This application is the national phase of international application PCT/GB98/02259 filed July 28, 1998 which designated the U.S.--
b. ☐ --This application also claims the benefit of U.S. Provisional Application No. 60/ _____, filed _____ .--

12. ☐ Amendments to the claims of the International Application **under PCT Article 19 (35 U.S.C. 371(c)(3)), i.e., before 18th month from first priority date above in item 3, are transmitted herewith (file only if in English) including:**

13. ☒ PCT Article 19 claim amendments (if any) have been transmitted by the International Bureau

14. ☐ Translation of the amendments to the claims **under PCT Article 19 (35 U.S.C. 371(c)(3)), i.e., of claim amendments** made before 18th month, **is attached (required by 20th month from the date in item 3 if box 4(a) above is X'd, or 30th month if box 4(b) is X'd, or else amendments will be considered canceled).**

15. **A declaration of the inventor** (35 U.S.C. 371(c)(4))
a. ☒ is submitted herewith ☒ Original ☐ Facsimile/Copy
b. ☐ is not herewith, but will be filed when required by the forthcoming PTO Missing Requirements Notice per Rule 494(c) if box 4(a) is X'd or Rule 495(c) if box 4(b) is X'd.

16. **An International Search Report (ISR):**
a. Was prepared by ☒ European Patent Office ☐ Japanese Patent Office ☐ Other
b. ☒ has been transmitted by the international Bureau to PTO.
c. ☒ copy herewith (2 pg(s).) ☒ plus Annex of family members (1 pg(s).).

17. **International Preliminary Examination Report** (IPER):
a. ☒ has been transmitted (if this letter is filed after 28 months from date in item 3) in English by the International Bureau with Annexes (if any) in original language.
b. ☒ copy herewith in English.
c.1 ☐ IPER Annex(es) in original language ("Annexes" are amendments made to claims/spec/drawings during Examination) including attached amended:
c.2 ☐ Specification/claim pages #_____ claims #
Dwg Sheets #
d. ☐ Translation of Annex(es) to IPER **(required by 30th month due date, or else annexed amendments will be considered canceled).**

18. **Information Disclosure Statement** including:
a. ☒ Attached Form PTO-1449 listing documents
b. ☐ Attached copies of documents listed on Form PTO-1449
c. ☒ A concise explanation of relevance of ISR references is given in the ISR.

19. ☒ **Assignment** document and Cover Sheet for recording are attached. Please mail the recorded assignment document back to the person whose signature, name and address appear at the end of this letter.

20. ☐ Copy of Power to IA agent.

21. ☐ **Drawings** (complete only if 8d or 10a(4) not completed): ___ sheet(s) per set: ☐ 1 set informal; ☐ Formal of size ☐ A4 ☐ 11"

22. ☐ _____ (No.) **Verified Statement(s)** establishing "small entity" status under Rules 9 & 27

23. **Priority** is hereby claimed under 35 U.S.C. 119/365 based on the priority claim and the certified copy, both filed in the International Application during the international stage based on the filing in (country) UNITED KINGDOM of:

	Application No.	Filing Date		Application No.	Filing Date
(1)	9716162.4	August 1, 1997	(2)		
(3)			(4)		
(5)			(6)		

a. ☒ See Form PCT/IB/304 sent to US/DO with copy of priority documents. If copy has not been received, please proceed promptly to obtain same from the IB.
b. ☐ Copy of Form PCT/IB/304 attached.

RE: USA National Filing of PCT/GB98/02259

420 Rec'd PCT/PTO 01 FEB 2000

24. Attached: 16 pages of Sequence Listing

25. **Preliminary Amendment:**

25.5 Per Item 17.c2, **cancel original** pages #_____, claims #_____, Drawing Sheets #

26. **Calculation of the U.S. National Fee (35 U.S.C. 371 (c)(1)) and other fees is as follows:**
Based on amended claim(s) per above item(s) ☐ 12, ☐ 14, ☐ 17, ☐ 25, ☐ 25.5 (hilite)

Total Effective Claims	22	minus 20 =	2	x $18/$9 =	$36	966/967
Independent Claims	6	minus 3 =	3	x $78/$39 =	$234	964/965
If any proper (ignore improper) Multiple Dependent claim is present,				add$260/$130	+260	968/969

BASIC NATIONAL FEE (37 CFR 1.492(a)(1)-(4)): →→ **BASIC FEE REQUIRED, NOW** →→→→

A. If country code letters in item 1 are **not** "US","BR","BB","TT","MX","IL" "NZ", "IN" or "ZA"

See item 16 re:

1. Search Report was not prepared by EPO or JPO - - - - - - - - -	add$970/$485		960/961
2. Search Report was prepared by EPO or JPO - - - - - - - - - - - -	add$840/$420	+840	970/971

SKIP B, C, D AND E UNLESS country code letters in item 1 are "US","BR","BB","TT","MX","IL", "NZ", "IN" or "ZA"

→ (X) ☐	B. If USPTO did not issue both International Search Report (ISR) and (if box 4(b) above is X'd) the International Examination Report (IPER), -	add$970/$485	+0	960/961
(only) (one) → (of) ☐	C. If USPTO issued ISR but not IPER (or box 4(a) above is X'd), -	add$690/$345	+0	958/959
(these) (4) → (boxes) ☐	D. If USPTO issued IPER but IPER Sec. V boxes not all 3 YES, -	add$670/$335	+0	956/957
→ ☐	E. If international preliminary examination fee was paid to USPTO and Rules 492(a)(4) and 496(b) satisfied (IPER Sec. V all 3 boxes YES for all claims), - - - - - - - - - - - - -	add $96/$48	+0	962/963

27. **SUBTOTAL =** $1370

28. If Assignment box 19 above is X'd, add Assignment Recording fee of ----$40 +40 (581)

29. Attached is a check to cover the - **TOTAL FEES** $0

Our Deposit Account No. 03-3975
Our Order No. 9901 (C#) | 265461 (M#)

CHARGE STATEMENT: The Commissioner is hereby authorized to charge any fee specifically authorized hereafter, or any missing or insufficient fee(s) filed, or asserted to be filed, or which should have been filed herewith or concerning any paper filed hereafter, and which may be required under Rules 16-18 and 492 (missing or insufficient fee only) now or hereafter relative to this application and the resulting Official document under Rule 20, or credit any overpayment, to our Account/Order Nos. shown above for which purpose a duplicate copy of this sheet is attached.
This CHARGE STATEMENT does not authorize charge of the issue fee until/unless an issue fee transmittal form is filed

Pillsbury Madison & Sutro LLP
Intellectual Property Group

1100 New York Avenue, NW
Ninth Floor
Washington, DC 20005-3918
Tel: (202) 861-3000
Atty/Sec: DJB/mhn

By Atty: Donald J. Bird **Reg. No.** 25323
Sig: _____ **Fax:** (202) 822-0944
Tel: (202) 861-3027

NOTE: File in duplicate with 2 postcard receipts (PAT-103) & attachments.

ZGGBP1, NOVEL PEPTIDES RELATED TO BIPOLAR AFFECTIVE DISORDER TYPE 1, SEQUENCES AND USES THEREOF

This invention relates to a novel human gene (ZGGBP1) associated with affective neurological disorders such as bipolar affective disorder. The invention also relates to homologues of the ZGGBP1 gene in species such as rat and mouse useful in providing animal models of affective disorders. The invention further relates to both the cDNA and the structural gene and to fragments encoding functional domains within the gene. The invention also relates to means for producing the protein encoded by the gene and to means for regulating its production and activity in vivo.

Affective disorders comprise a broad and heterogeneous category of psychiatric illness with a prevalence of up to 20% in the population. The most severe of these disorders is bipolar type I which affects approximately 1% of the population and this rate is fairly consistent across countries. The disease affects young adults, with a mean age of onset of 22 years. Treatment depends upon the phase of the disease and pharmacological agents include lithium carbonate, carbamazepine or valproic acid, tricyclic antidepressants. Monoamine oxidase inhibitors and selective serotonin re-uptake inhibitors are now also being used. The success rate of individual drugs is variable and some patients are treated with a combination of agents, although most have some unwanted side-effects. At present the precise diagnosis of individual affective disorders is difficult and new, gene based, diagnostic methods are desirable.

Family, twin and adoption studies have suggested the importance of genetic predisposition to bipolar affective disorder. On this basis, several groups have undertaken genetic linkage analysis in families with a high incidence of the disorder to find a causal gene. Many of the studies show conflicting data suggesting that a single gene is unlikely to be the cause. Rather, multiple interacting genetic traits may be involved. A recent study (Stine et al. 1995) identified two regions on chromosome 18 showing linkage to the disease.

The present invention is based on our discovery of a novel gene which maps to 18q21 and which unexpectedly shows appreciable sequence homology to the ned-4 gene on chromosome 15. Ned-4 is the human homologue of the mouse nedd-4 gene which is known to be differentially expressed during neural development and to be involved in signal transduction. Human ned-4 has been shown (Schild et al. 1996, Straub

et al. 1996) to be a negative regulator of a sodium channel which is deleted in Liddle's syndrome (a hereditary form of hypertension).

Nedd-4 was originally isolated as a partial cDNA clone from a mouse brain library (Kumar et al. 1992) as one of a set of genes which were differentially expressed during development (Neural precursor cells expressed developmentally down-regulated). The derived amino acid sequence contains three copies of the WW domain (Andre & Springael 1994, Bork & Sudol, 1994; Hofmann & Boucher, 1995), a Ca lipid binding (CaLB/C2) domain (Brose et al. 1995) and a Hect (homologous to the E6-AP carbodyl terminus) domain which has homology to a ubiquitin ligase (E3) enzyme (Huibregtse et al. 1995). The human homologue of nedd-4 (Ned-4) was isolated as an randomly cloned EST (KIAA0093) from immature myeloblast mRNA (Nomura et al. 1994) and shown by sequence comparison to have 86% identity at the amino acid level to the mouse sequence. The human sequence, however, has a fourth copy of the WW domain.

The WW domain is a 40 amino acid sequence found in several unrelated proteins. The two highly conserved tryptophans give it its name. The function of the domain is thought to be involved in protein-protein interactions. Despite their functional diversity, the proteins listed all appear to be involved in cell signalling or regulation. It has been shown that the WW domains of Nedd-4 interact with the proline-rich PY motifs in the epithelial sodium channel in the kidney (Schild et al. 1996). Mutational deletion of the PY motifs in the epithelium sodium channel in Liddle's syndrome, an inherited disease causing systemic hypertension characterised by hyperactivity of the sodium channel, has been shown to abrogate binding of Nedd-4 (Straub et al. 1996). It is therefore likely that Nedd-4 has a negative regulatory role when bound to the channel.

The Hect domain is an E3 ubiquitin-protein ligase domain and enzymes with this domain catalyse polyubiquitination, which is involved in several cellular processes including proteolytic degradation.

The CaLB/C2 domain is thought to be involved in calcium-dependent phospholipid binding, although some proteins containing this domain do not bind calcium and other putative functions for the C2 domain such as binding to inositol -1,3,4,5-tetraphosphate have been suggested. Examples of proteins containing this domain are Protein Kinase C (PKC) isoenzymes and synaptogamins.

PCT patent application WO97/12962 discloses a protein (Pub3) with homology to Pub1, a Schizosaccaromyces Pombe protein which has an apparent function in the ubiquitination of, among other cellular proteins, the mitotic activating tyrosine phosphatase cdc25 and the tumour suppresser protein p53. As such this protein may be involved in regulating the progression of proliferation in eukaryotic cells by effectively controlling the activity of the cdk complexes by modulating the availability of cdc25 and/or p53.

A comparison of Pub3 with ZGGBP1 revealed that the sequences represent two distinct genes which code for two separate, structurally unrelated proteins. The two genes share sequence homology within a certain defined region, the sequences are identical within the region 516-3568 of ZGGBP1, but they do not show any homology within the regions 5' and 3'of this sequence. In addition the derived amino acid sequence for ZGGBP1 is completely different to that derived for Pub 3 as both have been initiated from a different start methionine. A comparison of the nucleotide sequences for ZGGBP1 and Pub 3 is outlined in Figure 5.

Therefore in a first aspect of the present invention we provide the ZGGBP1 gene having the full length cDNA as set out in SEQ ID NO: 1. We further provide fragments of the ZGGBP1 gene comprising ZGGBP1 sequence outside the region defined by base pairs 516-3568 of the ZGGBP1 gene. By fragments we mean contiguous regions of the gene including complementary DNA and RNA sequences, starting with short sequences useful as probes or primers of say about 8-50 bases, such as 10-30 bases or 15-35 bases, to longer sequences of up to 50, 100, 200, 500 or 1000 bases. Indeed any convenient fragment of the gene of say up to 2kb, 3kb, 4kb or more than 4kb may be a useful gene fragment for further research, therapeutic or diagnostic purposes. Further convenient fragments include those whose terminii are defined by restriction sites within the gene of one or more kinds, such as any combination of Rsa1, Alu1 and Hinf1.

In a further aspect of the invention we provide homologues of the ZGGBP1 gene in species such as rat and mouse useful in providing animal models of affective disorders. By homologue, we mean a corresponding ZGGBP1 gene in another species, which displays greater than 85% sequence homology, conveniently greater than 90%, for example 95%, to the human ZGGBP1 sequence. The full sequences of the individual homologues may be determined using conventional techniques such as hybridisation, PCR

and sequencing techniques, starting with any convenient part of the sequence set out in SEQ ID NO: 1. The partial sequence of the mouse gene is set out in SEQ ID NO: 3 and this gene and the protein encoded by this gene represent further independent aspects of the invention.

In a further aspect of the invention we provide polynucleotide sequences capable of specifically hybridising to the ZGGBP1 gene. By specifically hybridising we mean that the polynucleotide hybridises under stringent conditions to the sequence on chromosome 18q21 as set out in SEQ ID No: 1, or to the corresponding non-coding sequence, to the exclusion of other genomic loci. It is contemplated that a species such as a peptide nucleic acid may be an acceptable equivalent to a polynucleotide, at least for purposes that do not require translation into protein.

In a further aspect of the invention we provide a recombinant ZGGBP1 protein obtained by expression of all or a part of the cDNA as set out in SEQ ID NO: 1. The recombinant protein may comprise all or a convenient part of the peptide sequence set out in SEQ ID NO: 2. The production of a protein according to the invention may be achieved using standard recombinant DNA techniques involving the expression of the protein by a host cell as described for example by Sambrook et al. 1989. The isolated nucleic acids described herein may for example be introduced into any convenient expression vector for example the T7 Studier system for expression in E.coli (US-A-4952496), Pichia pastoris for expression in yeast, the Baculovirus system for expression in insect cells and the GS system for expression in mammalian cells by operatively linking the DNA to any necessary expression control elements therein and transforming any suitable prokaryotic or eukaryotic host cell with the vector using well known procedures.

Therefore in a further aspect of the invention we provide a recombinant plasmid comprising all or a part of the ZGGBP1 cDNA of the invention.

The invention further extends to cells containing said recombinant plasmids and to a process for producing a ZGGBP1 protein of the invention which comprises culturing said cells such that the desired protein is expressed and recovering the protein from the culture.

By way of example, the nucleotide sequence in SEQ ID NO: 1 is inserted downstream of the SV40 promoter in the pGEX plasmid vector, and either transiently or

stably expressed in COS -7 cells. Expression of the protein according to the invention can be detected following disruption of the cells by Western blotting .

It may be desirable to produce the individual functional domains of the protein according to the invention in isolation from the rest of the molecule. This may be achieved using the above standard recombination DNA techniques except that in this instance the DNA sequence used is that encoding one of the partial amino acid sequences of the domains identified in Figure 1 or a combination of these.

By way of further example, the nucleotide sequence in SEQ ID NO: 1 is inserted downstream of the SV40 promoter and the glutathione-S-transferase (GST) coding sequence in the pBC plasmid vector, and either transiently or stably expressed in COS -7 cells allowing expression of the corresponding fusion protein. Expression of the fusion protein can be detected following disruption of the cells by Western blotting with antibodies to GST, and furthermore the fusion protein can be used in an affinity binding procedure to find proteins which are functional partners of the protein of the invention from cell extracts.

A ZGGBP1 protein of the invention may in particular be used to screen for compounds which regulate the activity of the enzymes and the invention extends to such a screen and to the use of compounds obtainable therefrom to regulate the activity of the protein in vivo.

Thus according to a further aspect of the invention we provide a method for identifying a compound capable of modulating the action of a ZGGBP1 protein which method comprises subjecting one or more test compounds to a screen comprising (A) a protein containing the amino acid sequence shown in SEQ ID NO: 2 or a homologue or fragment thereof, or (B) the nucleotide sequence shown in SEQ ID NO: 1 or a homologue or fragment thereof, or (C) a host cell expressing a ZGGBP1 polypeptide or a homologue or fragment thereof.

The screen according to the invention may be operated using conventional procedures, for example by bringing the test compound or compounds to be screened and an appropriate substrate into contact with the protein or a cell capable of producing it and determining affinity for the protein in accordance with conventional procedures.

Any compound identified in this way may be used in the treatment of humans and/or other animals of one or more of the above mentioned diseases. The invention thus

extends to a compound selected through its ability to regulate the activity of the protein in vivo as primarily determined in a screening assay utilising the protein containing an amino acid sequence shown in SEQ ID NO: 2 or a homologue or fragment thereof, or a gene coding therefor for use in the treatment of a disease in which the over- or under-activity or unregulated activity of the protein is implicated.

In a further aspect of the invention we provide examples of insertions/deletions and single base change polymorphisms (mutations) as outlined in Figure 6, 7, 8, 9 and 10.

The ZGGBP1 gene of the invention may also be used as the basis for diagnosis, for example to determine expression levels in a human subject, by for example direct DNA sequence comparison or DNA/RNA hybridisation assays. Diagnostic assays may involve the use of nucleic acid amplification technology such as the PCR and in particular the Amplification Refractory Mutation System (ARMS) as claimed in our European Patent No. 0 332 435. Such assays may be used to determine allelic variants of the gene, for example insertions, deletions and/or mutations such as one or more point mutations. Such variants may be heterozygous or homozygous.

In a further aspect of the invention, amplification primers may be provided for use in the above diagnostic methods. In general, these are provided as a set and used for PCR amplification. One of the primers conveniently hybridises to a ZGGBP1 locus outside the region defined by base pairs 516-3568 thus allowing the ZGGBP1 gene on 18q21 to be identified to the exclusion of other loci.

The ZGGBP1 gene may also be used in gene therapy, for example where it is desired to modify the production of the protein in vivo, and the invention extends to such uses.

Knowledge of the gene according to the invention also provides the ability to regulate its expression in vivo by for example the use of antisense DNA or RNA. Thus, according to a further aspect of the invention we provide an antisense DNA or an antisense RNA which is complementary to the polynucleotide sequence shown in SEQ ID NO: 1. By complementary we mean that the two molecules can base pair to form a double stranded molecule.

The antisense DNA or RNA for co-operation with the gene in SEQ ID NO: 1 can be produced using conventional means, by standard molecular biology and/or by chemical synthesis as described above. If desired, the antisense DNA or antisense RNA may be

chemically modified so as to prevent degradation in vivo or to facilitate passage through a cell membrane and/or a substance capable of inactivating mRNA, for example ribozyme, may be linked thereto and the invention extends to such constructs.

The antisense DNA or antisense RNA may be of use in the treatment of diseases or disorders in humans in which the over- or under-regulated production of the gene product has been implicated. Such diseases or disorders may include those described under the general headings of neurologic, eg.stroke, dementia, renal eg. hypertension, nephrosis, cardiovascular disorders.

Convenient DNA sequences may be obtained using conventional molecular biology procedures, for example by probing a human genomic or cDNA library with one or more labelled oligonucleotide probes containing 10 or more contiguous nucleotides designed using the nucleotide sequences described here. Alternatively, pairs of oligonucleotides one of which is homologous to the sense strand and one to the antisense strand, designed using the nucleotide sequences described here to flank a specific region of DNA may be used to amplify that DNA from a cDNA library.

The ZGGBP1 protein of the invention and homologues or fragments thereof may be used to generate substances which selectively bind to it and in so doing regulate the activity of the protein. Such substances include, for example, antibodies, and the invention extends in particular to an antibody which is capable of recognising one or more epitopes containing the protein binding domains shown in Figure 1. In particular the antibody may be neutralising antibody.

As used herein the term antibody is to be understood to mean a whole antibody or a fragment thereof, for example a F(ab)2, Fab, FV,. VH or VK fragment, a single chain antibody, a multimeric monospecific antibody or fragment thereof, or a bi- or multi-specific antibody or fragment thereof.

The invention will now be illustrated but not limited by reference to the following detailed description, References, Examples and Figures wherein:

Figure 1 shows the predicted amino acid sequence of ZGGBP1. The C2 domain is indicated by carets , the four WW domains are indicated by asterisks and the Hect domain is indicated by underlining .

Figure 2 shows a comparison of amino acid sequences of human ned4 Swissprot entry P46934 and ZGGBP1.

Figure 3 shows a Northern blot analysis of various human tissues probed with ZGGBP1.

Figure 4 shows a comparison of the nucleic acid sequences of human and mouse ZZGBP1. The mouse sequence is a partial cDNA which spans the C-terminal portion of the human protein coding region.

Figure 5 shows a comparison of the nucleic acid sequences for ZGGBP1 and Pub3

Figure 6 shows a polymorphism located at position 3554 of the cDNA sequence

Figure 7 shows a polymorphism located at position 4828 of the cDNA sequence

Figure 8 shows a polymorphism located in an intronic sequence derived from a BAC containing ZGGBP1

Figure 9 shows a variable number of tetranucleotide repeats located within an intronic sequence from ZGGBP1

Figure 10 shows an insertion at position 4032 of the cDNA sequence

References

Stine O.C. et al. (1994) Am.J.Hum.Genet. *57*, 1384-1394;

Lovett M., Kere J. and Hinton L. (1991) Proc.Natl.Acad.Sci. USA *88*, 9628-9632;

Schuler G.D. et al. (1996) Science *274* , 540-546;

Kumar S., Tomooka Y. and Noda M. (1992) Biochem.Biophys.Res.Commun. *185*, 1155-1161;

Andre B. and Springael J.Y. (1994) Biochem.Biophys.Res.Commun. *205*, 1201-1205;

Bork P. and Sudol M. (1994) Trends Biochem.Sci. *19*, 531-533;

Hofmann K. and Boucher P. (1995) FEBS Letts. *358*, 153-157;

Brose N., Hofmann K.O., Hata Y., Suedhof T.C.(1995) J. Biol. Chem. *270*, 25273-25280;

Huibregtse J.M., Scheffner M., Beaudenon S. and Howley P.M. (1995) Proc.Natl.Acad.Sci. USA, *92*, 2563-2567;

Nomura N. et al. (1994) DNA Res. *2*, 37-43;

Schild L., Lu Y., Gautschi I., Schneeberger E., Lifton R.P. and Rossier B.C. (1996) EMBO J. *15*, 2381-2387;

Straub O., Dho S., Henry P.C., Correa J., Ishikawa T., McGlade J. and Rotin D. (1996) EMBO J. *15* , 2371-2380;

Shizua H., Birren B., Kim U-J, Mancino V., Slepak T., Tachiiri Y. and Simon M. (1992) Proc.Natl.Acad.Sci. USA *89*, 8794-8797;

Kim U-J., Shizuya H., Kang H-L., Choi S-S., Garrett C.L., Smink L.J., Birren B.W., Korenberg J.R., Dunham I. and Simon M. (1996) Proc.Natl.Acad.Sci. USA *93*, 6297-6301.

Sambrook J., Fritsch E.F. and Maniatis T. (1989) "Molecular Cloning: A Laboratory Manual (2nd edition)" Cold Spring Harbor Laboratory Press NY

Example 1

Identification of ZGGBP1

We used two methods for investigating the 18q21 region of interest. In one method we used positional cloning to identify novel transcripts from physical clones representing the region and in a second method we utilised public databases to identify transcripts which had been assigned to a low resolution map of the region by radiation hybrid mapping and assigned them to physical clones representing a high resolution map of the region.

Method 1 - Positional Cloning

The 18q21 region described by Stine et al. (1995) is delimited by the STS markers used by that group to identify linkage. They found the most strongly linked marker to be D18S41, which had a LOD score of 3.51 in cases of paternal inheritance. Linkage declined over flanking markers. We identified a set of four Yeast Artificial Chromosomes (YACs) which comprised a contiguous overlapping set of genomic clones covering the defined region by the presence in those YACs of STS markers used in the Stine study.

DNA from the YACs was prepared and used in a PCR-based hybridisation approach to enrich for transcripts from a human fetal brain cDNA library. This approach, known as direct selection (Lovett et al. 1991) has been shown to be efficient in identifying transcripts present on large genomic clones.

Method 2 - Refining Radiation Hybrid Mapped Transcripts

The UNIGENE database is a repository for transcripts which have been mapped by taking representative Expressed Sequence Tagged Sites (ESTs) and performing PCR analysis on a panel of radiation hybrids which have been calibrated with respect to a framework of 1000 genetic markers (Schuler et al. 1996). We found 36 EST clusters which had been mapped to a radiation hybrid map interval which corresponded to the 18q21 region of interest and to flanking regions outside.

All the ESTs were tested by PCR on our YAC genomic clones to determine which were present. We found approximately half of the ESTs to be present within the genomic clones and were able to order them based on their position within the YAC contig.

Results

Several clones from our direct selection experiments showed sequence homology to a known EST which we had previously shown to be present in two of the YACs within the contig. The EST was representative of a cluster of sequences. All of these sequences were assembled together using DNAStar Seqman and the consensus sequences obtained were used iteratively to search for other database members within both Unigene, dbEST and EMBL databases. This resulted in the surprising identification of two further clusters of ESTs which had previously not been related to each other on the basis of sequence analysis. The two new EST clusters were annotated as having sequence similarity to ned-4. This was an unexpected finding since we had recently mapped the human ned-4 by Fluorescence In Situ Hybridisation (FISH) to chromosome 15. We were aware that ned-4 was involved in neuronal cell signalling and we concluded that the EST cluster on 18q21 must represent a closely related gene and therefore likely to be involved in affective neurological disorders such as bipolar affective disorder.

The assembly of the EST clusters did not give rise to a single complete contiguous sequence. The reason for this is that many of the EST sequences were derived from IMAGE cDNA clones for which end sequence only was available. In order to fill in the gaps and give a complete contig, four of these clones (IMAGE I.D. 80951, 33059, 79526 and 79984) were sequenced completely to fill the gaps and give an entire complete contiguous sequence. Comparison of the sequence with ned-4 showed that the contig comprised 2kb of 3'Untranslated Region (UTR) and 700bp of the coding region of a gene

which had approximately 85% identity at the amino acid level to ned-4 and which we named ZGGBP1.

Isolation of the full length gene for ZGGBP1

The extending of partial transcripts to full length clones can be a complex and difficult process requiring skill and expertise for success. Having considered several possibilities, we opted for a PCR-based approach to isolate and characterise the full length ZGGBP1 gene. Human foetal brain double stranded cDNA was synthesised from mRNA using standard methods (Sambrook et al. 1989) and ligated into lambda Zap vector by use of adapters. However, in order to minimise the loss of transcripts often seen following the cloning step, the resulting ligation mix was not cloned but was instead used as a template for PCR. Oligonucleotide primers specific to ZGGBP1 were used in combination with vector specific primers to amplify DNA across the unknown part of the gene. Since the distance to be covered was unknown, we performed long PCR using the commercially available BCL Expand enzyme and long (30mer) oligonucleotide primers. Since we were using unamplified material, where our target cDNAs were likely to be present only in very small amounts, we utilised a secondary PCR step with nested oligonucleotide primers and again using long PCR to yield sufficient PCR products to be visible by gel analysis and also to minimise the possibility of non-specific PCR amplification. The PCR products derived from these experiments were then purified and sequenced directly. Where necessary, the DNA sequence obtained was used to design further primers to walk along the gene in a 3' - 5' direction. The complete nucleotide sequence derived from this work is 5.2kb and the translated amino acid sequence is shown in SEQ ID NO: 1.

The amino acid sequence derived from the cDNA was compared with that of ned-4 and is shown in Figure 2. The proteins diverge markedly towards the N-terminal portion of the protein, although there is conservation of the common functional motifs.

Northern analysis using a probe derived from the 3'UTR of ZGGBP1 showed a band at approximately 4.8kb but also a more abundant band of 9kb in size in several neurological tissues, with the exception of medulla or spinal cord. These bands are likely to be due to alternative splicing (Figure 3). Other tissues contained the 4.8kb band at higher abundance with respect to the 9kb band and also a 4kb band. ZGGBP1 was

expressed in all tissues examined with the exception of liver where we could not detect a transcript at our current detection sensitivity.

Comparison of Amino Acid Sequences of human ned-4 and ZGGBP1

A comparison of the amino acid sequences of human ned-4 and ZGGBP1 is shown in Figure 6. The two proteins have a high level of homology over much of the C-terminal region, including the Hect and WW domains, but diverge over the central portion of the protein. There is a further block of homology near to the N-terminal region, including the C2 domain. The presence of these domains in ZGGBP1 suggests some common functionality with ned-4.

Identification of polymorphic variants of ZGGBP1

500bp regions of the ZGGBP1 cDNA were PCR amplified from a variety of tissues and lymphoblastoid cell lines. Sequencing was carried out and polymorphisms identified as outlined in Figures 5 and 6. Some intronic sequence had been identified from a genomic clone and sequence analysis of these regions identified a further polymorphic variant as outlined in Figure 7. A tetranucleotide repeat (GATT) was also identified in an intronic sequence derived from this BAC and this was found to have variable numbers of repeats (Figure 8).

Isolation of Genomic Clone for ZGGBP1

The Research Genetics human Bacterial Artificial Chromosome (BAC) library (Shizua et al. 1992, Kim et al. 1996) was screened by PCR using primers specific to the 3'UTR of ZGGBP1 and BACs were isolated. These are being used to characterise the structural gene including the intron/exon structure and the 5' regulatory region.

Isolation of Mouse homologue for ZGGBP1

The full length sequence of ZGGBP1 shown in SEQ ID NO: 1 was used to search the dbEST database to identify homologous mouse sequences. Three overlapping IMAGE clones were identified (IMAGE I.D.479436, 573510, 482922) comprising a partial transcript. Comparison of the mouse and human nucleotide sequence is shown in Figure 4. The mouse clones were isolated for use as a probe for in situ hybridisation on sections

of mouse brain during development, and as a probe of mouse genomic libraries to isolate genomic clones and to produce transgenic mice by gene targeting using homologous recombination.

CLAIMS

1. A polynucleotide comprising a nucleic acid sequence which encodes the polypeptide of Seq ID No 2, and homologues and fragments thereof.

2. A polynucleotide as claimed in claim 1 which comprises the cDNA sequence of Seq ID No 1.

3. Polymorphic variants of the polynucleotide as claimed in claim 2, selected from the group in which:

i) T at position 3554 is replaced by C.

ii) C at position 4828 is replaced by G.

iii) T within an intronic region associated with ZGGBP1 is replaced by C.

iv) C is inserted at position 4032.

4. A polynucleotide which comprises an animal homologue of the nucleic acid claimed in claims 1-3.

5. A polynucleotide as claimed in claim 4 which comprises the cDNA sequence of Seq ID No 3, and homologues and fragments thereof.

6. A polynucleotide which is capable of specifically hybridising to eight or more contiguous nucleotides comprised in Seq ID No 1 or Seq ID No 3 or comprised in the complementary strands thereof.

7. A polynucleotide which comprises a ZGGBP1 gene fragment.

8. A vector comprising a polynucleotide of claims 1-7.

9. A host cell transformed with a vector of claim 8.

10. A polypeptide comprising the amino acid sequence of Seq ID No 2 and homologues and fragments thereof.

11. A polypeptide comprising the amino acid sequence of Seq ID No 4 and homologues and fragments thereof.

12. A fusion protein in which a polypeptide of claim 10 or claim 11 is fused with glutathione-S-transferase.

13. A method for producing cells which express a polypeptide of claim 10 or claim 11 or a fusion protein of claim 12, comprising:

a) culturing a host cell of claim 9 under conditions suitable for the expression of the polypeptide.

b) recovering the polypeptide from the host cell culture.

14. A method for identifying a compound capable of modulating the activity of a ZGGBP1 protein , which method comprises subjecting one or more test compounds to a screen comprising:

a) a protein as claimed in claims 10-12 or a homologue or fragment thereof,

or

b) a polynucleotide as claimed in claims 1-7 or a homologue or fragment thereof,

or

c) a host-cell expressing a polypeptide of a ZGGBP1 molecule,

and measuring an effect of the test compound on ZGGBP1 activity.

15. A compound that modulates the activity of a human ZGGBP1 identified by the method of claim 14.

16. A pharmaceutical composition comprising a compound that modulates the activity of a protein identified by the method of claim 14.

17. A diagnostic assay for the detection of ZGGBP1, which assay comprises measuring the presence or absence of a protein as claimed in claims 10-12 or a polynucleotide as claimed in claims 1-7.

18. An antisense molecule comprising a complement of the polynucleotide in claims 1-7 or a biologically effective fragment thereof.

19. Use of a polynucleotide as claimed in claims 1-7 or claim 18 in gene therapy.

20. An antibody specific for a protein of claims 10-12 or fragments thereof.

21. A set of amplification primers for selective amplification of a ZGGBP1 gene sequence.

FIGURE 1

```
MFRLRSWASSTTGSRYGSAFCGSPTLAWCVCVPVCYGESRILRVKVVSG

IDLAKKDIFGASDPYVKLSLYVADENRELALVQTKTIKKTLNPKWNEEF
                  ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^

YFRVNPSNHRLLFEVFDENRLTRDDFLGQVDVPLSHLPTEDPTMERPYT
^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^

FKDFLLRPRSHKSRVKGFLRLKMAYMPKNGGQDEENSDQRDDMEHGWEV

VDSNDSASQHQEELPPPPLPPGWEEKVDNLGRTYYVNHNNRTTQWHRPS
                        ********************************

LMDVSSESDNNIRQINQEAAHRRFRSRRHISEDLEPEPSEGGDVPEPWE
*
TISEEVNIAGDSLGVVLPPPPASPGSRTSPQELSEELSRRLQITPDSNG

EQFSSLIQREPSSRLRSCSVTDAVAEQGHLPPPSVAYVHTTPGLPSGWE
                                                ********

ERKDAKGRTYYVNHNNRTTTWTRPIMQLAEDGASGSATNSNNHLIEPQI
****************************

RRPRSLSSPTVTLXAPLEGAKDSPVRRAVKDTLSNPQSPQPSPYNSPKP

QHKVTQSFLPPGWEMRIAPNGRPFFIDHNTKTTTWEDPRLKFPVHMRSK

TSLNPNDLGPLPPGWEERIHLDGRTFYIDHNSKITQWEDPRLQNPAITG
          *********************************

PAVPYSREFKQKYDYFRKKLKKPADIPNRFEMKLHRNNIFEESYRRIMS

VKRPDVLKARLWIEFESEKGLDYGGVAREWFFLLSKEMFNPYYGLFEYS

ATDNYTLQINPNSGLCNEDHLSYFTFIGRVAGLAVFHGKLLDGFFIRPF

YKMMLGKQITLNDMESVDSEYYNSLKWILENDPTELDLMFCIDEENFGQ

TYQVDLKPNGSEIMVTNENKREYIDLVIQWRFVNRVQKQMNAFLEGFTE

LLPIDLIKIFDENELELLMCGLGDVDVNDWRQHSIYKNGYCPNHPVIQW

FWKAVLLMDAEKRIRLLQFVTGTSRVPMNGFAELYGSNGPQLFTIEQWG

SPEKLPRAHTCFNRLDLPPYETFEDLREKLLMAVENAQGFEGVD.
```

FIGURE 2

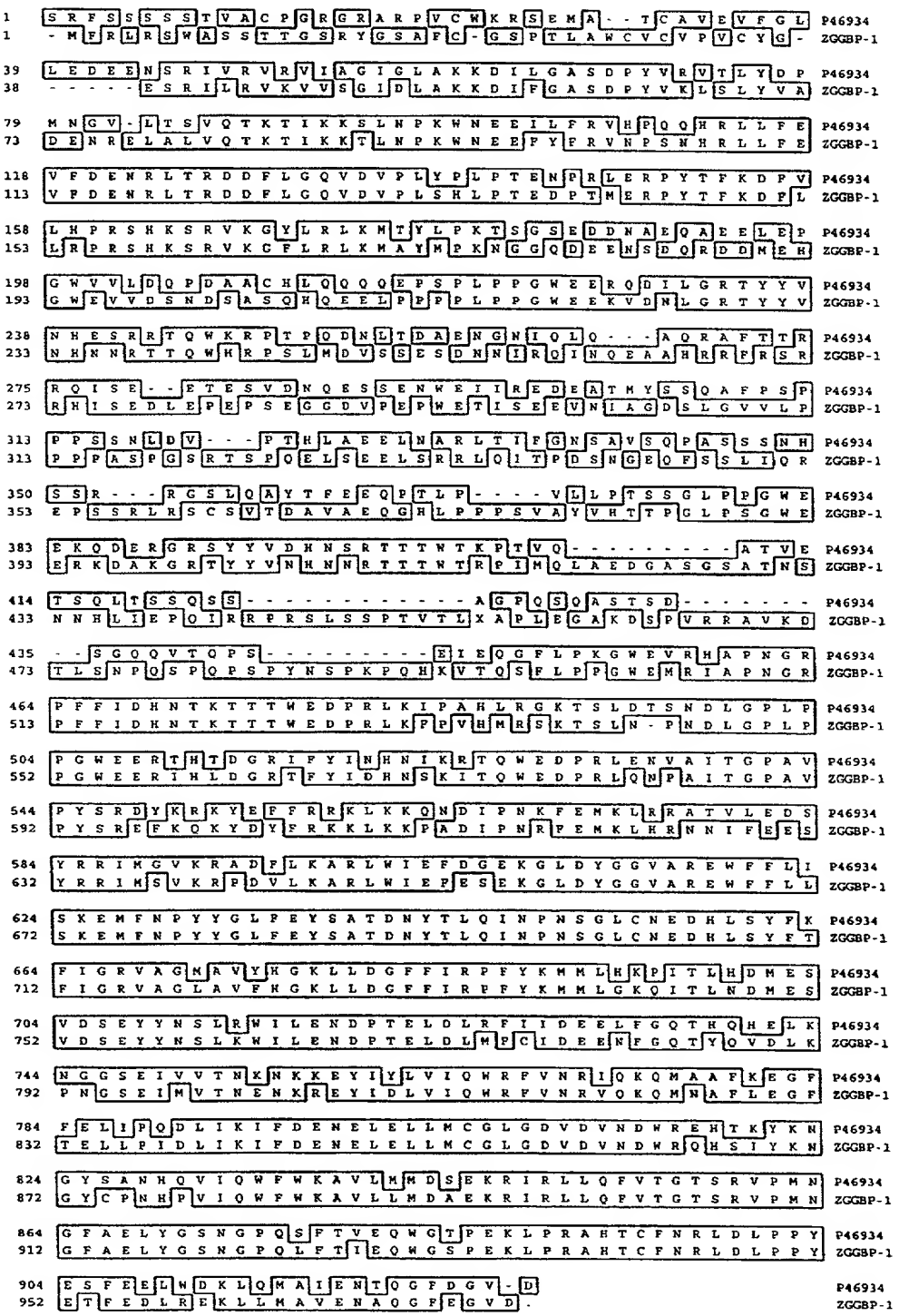


FIGURE 3

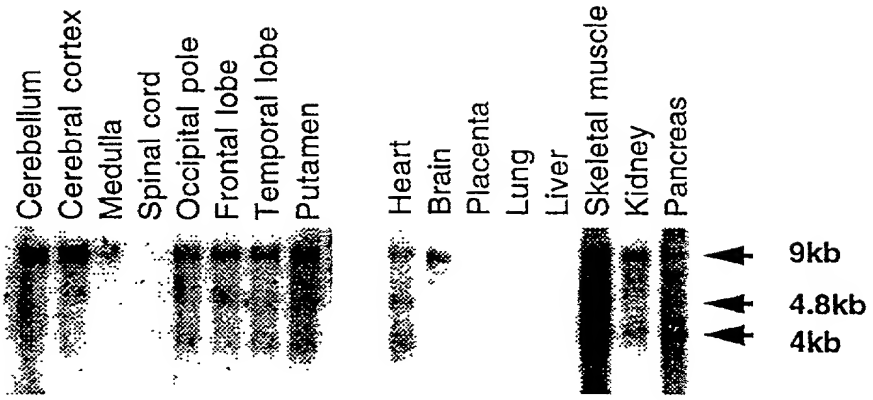

Cerebellum
Cerebral cortex
Medulla
Spinal cord
Occipital pole
Frontal lobe
Temporal lobe
Putamen
Heart
Brain
Placenta
Lung
Liver
Skeletal muscle
Kidney
Pancreas
9kb
4.8kb
4kb

FIGURE 4

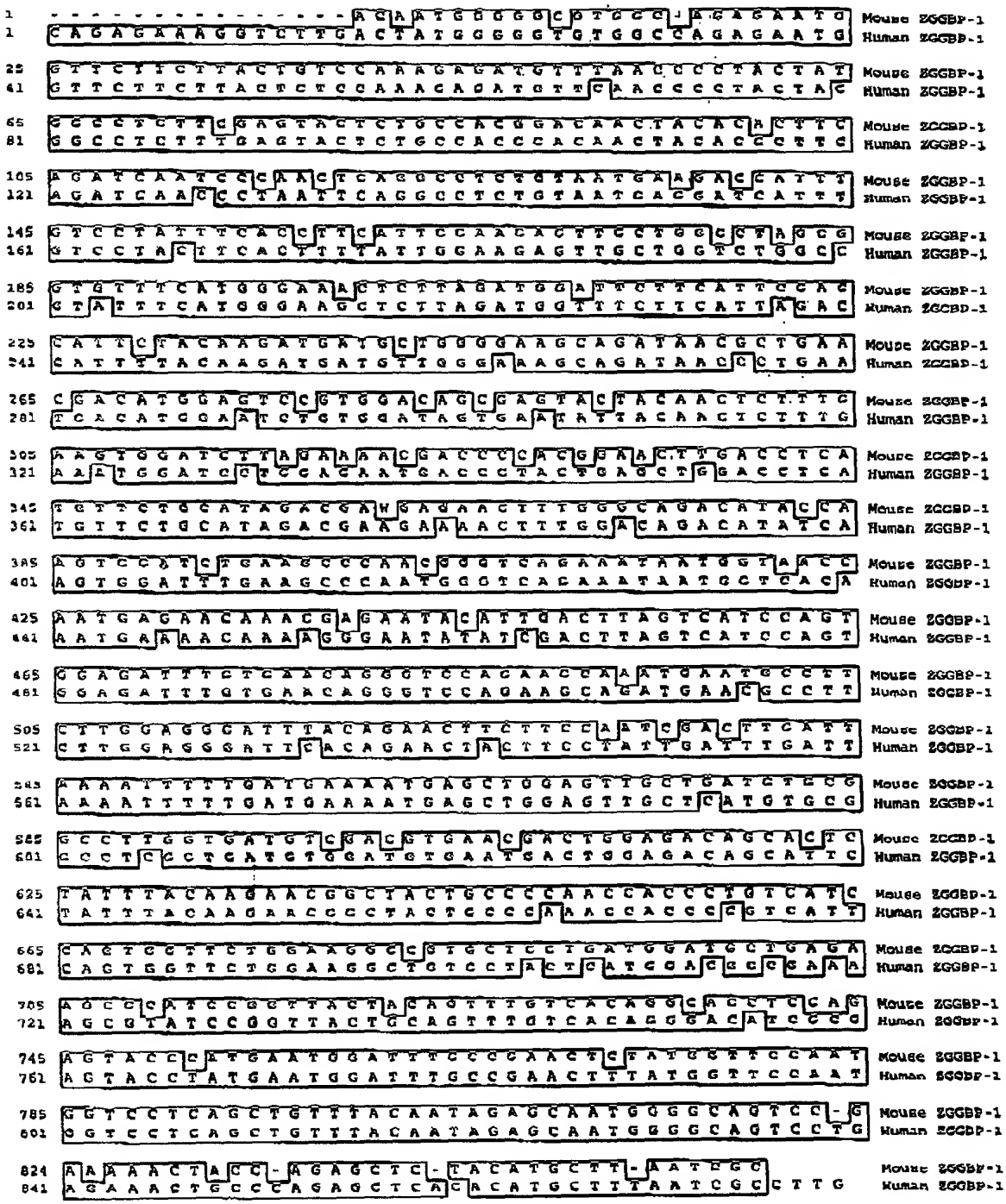


FIGURE 5

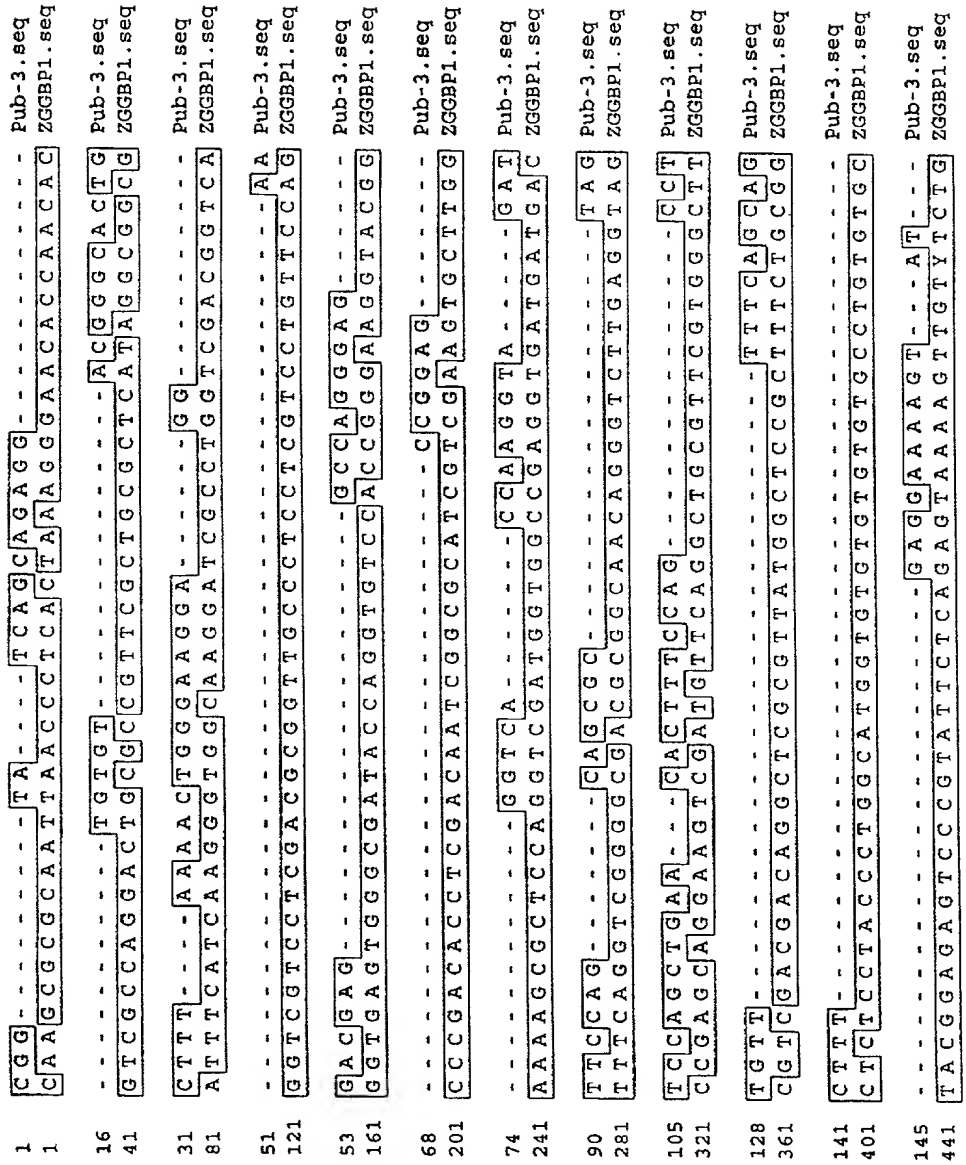

1 CGG----------TA----TCAGCAGAGG----------- Pub-3.seq
1 CAAGCGCGCAATTACCCTCACTAAGGGAACACCAACACAC ZGGBP1.seq
16 ----------TGTGT----------ACGGGCACTG Pub-3.seq
41 GTCGCCAGGACTGCGCTTCGCTGCGCTCATAGGCGCGGCG ZGGBP1.seq
31 CTTT----AAACTGGGAAGGA----------GG----------- Pub-3.seq
81 ATTTCATCAAGGGTGGCAAGGATCGCCTGGTCGACGGTCA ZGGBP1.seq
51 --A Pub-3.seq
121 GGTCGTCCTCGACGCGGTTGCCCTCCTCGTTCCTCAG ZGGBP1.seq
53 GACGAG----------GCCAGGGAG----------- Pub-3.seq
161 GGTGAGTGGCGAATACCAGGTGTCCACCGGGAGGTACGGG ZGGBP1.seq
68 ----------C GGGAG----------- Pub-3.seq
201 CCCGACACCTCGACAATCGGCGCATCGTCGACGGTTGG ZGGBP1.seq
74 ----------GGTCA----------CCAAGGGTA----GAT Pub-3.seq
241 AAAAGCGCTCCAGGTCGATGGTGGCCGAGGGTGATGATGAC ZGGBP1.seq
90 TTCCAG----------CAGCGC----------TAG Pub-3.seq
281 TTTCAGGTCGGGCGACGCGCAACACAGGGTCTTGAGGTAG ZGGBP1.seq
105 TCCAGCTGA----CACTTTCCAG----------CCT Pub-3.seq
321 CCGAGCAGGAAGTCGATGTTCAGGCTGCGTTCGTGGGCTT ZGGBP1.seq
128 TGTTT----------TTTCAGCAG Pub-3.seq
361 CGTCGACGACAGGCTCGCGTTATGGCTCCGCTTTCTGCGG ZGGBP1.seq
141 CTTTT----------GAGGAAAAGT----------- Pub-3.seq
401 CTCTCCTACCCTGGCATGGTGTTGTGTTGCCTGTGTTGC ZGGBP1.seq
145 ----------GAGGAAAAGT----------- Pub-3.seq
441 TACGGAGAGTCCGTATTCTCAGAGTTAAAAGTTGTYTCTG ZGGBP1.seq

FIGURE 5 continued

```
157  - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - A G T G A  Pub-3.seq
481  G A A T W G A T C T C G C C A A A A A G G A C A T C T T T G G A G C C A G T G A  ZGGBP1.seq

162  T C C G T A T G T G A A A C T T T C A T T G T A C G T A G C G G A T G A G A A T  Pub-3.seq
521  T C C G T A T G T G A A A C T T T C A T T G T A C G T A G C G G A T G A G A A T  ZGGBP1.seq

202  A G A G A A C T T G C T T T G G T C C A G A C A A A A A C A A T T A A A A A G A  Pub-3.seq
561  A G A G A A C T T G C T T T G G T C C A G A C A A A A A C A A T T A A A A A G A  ZGGBP1.seq

242  C A C T G A A C C C A A A A T G G A A T G A A G A A T T T T A T T T C A G G G T  Pub-3.seq
601  C A C T G A A C C C A A A A T G G A A T G A A G A A T T T T A T T T C A G G G T  ZGGBP1.seq

282  A A A C C C A T C T A A T C A C A G A C T C C T A T T T G A A G T A T T T G A C  Pub-3.seq
641  A A A C C C A T C T A A T C A C A G A C T C C T A T T T G A A G T A T T T G A C  ZGGBP1.seq

322  G A A A A T A G A C T G A C A C G A G A C G G C T T C C T G G G C C A G G T G G  Pub-3.seq
681  G A A A A T A G A C T G A C A C G A G A C G A C T T C C T G G G C C A G G T G G  ZGGBP1.seq

362  A C G T G C C C C T T A G T C A C C T T C C G A C A G A A G A T C C A A C C A T  Pub-3.seq
721  A C G T G C C C C T T A G T C A C C T T C C G A C A G A A G A T C C A A C C A T  ZGGBP1.seq

402  G G A G C G A C C C T A T A C A T T T A A G G A C T T T C T C C T C A G A C C A  Pub-3.seq
761  G G A G C G A C C C T A T A C A T T T A A G G A C T T T C T C C T C A G A C C A  ZGGBP1.seq

442  A G A A G T C A T A A G T C T C G A G T T A A G G G A T T T T T G C G A T T G A  Pub-3.seq
801  A G A A G T C A T A A G T C T C G A G T T A A G G G A T T T T T G C G A T T G A  ZGGBP1.seq

482  A A A T G G C C T A T A T G C C A A A A A A T G G A G G T C A A G A T G A A G A  Pub-3.seq
841  A A A T G G C C T A T A T G C C A A A A A A T G G A G G T C A A G A T G A A G A  ZGGBP1.seq

522  A A A C A G T G A C C A G A G G G A T G A C A T G G A G C A T G G A T G G G A A  Pub-3.seq
881  A A A C A G T G A C C A G A G G G A T G A C A T G G A G C A T G G A T G G G A A  ZGGBP1.seq

562  G T T G T T G A C T C A A A T G A C T C G G C T T C T C A G C A C C A A G A G G  Pub-3.seq
921  G T T G T T G A C T C A A A T G A C T C G G C T T C T C A G C A C C A A G A G G  ZGGBP1.seq
```

FIGURE 5 continued

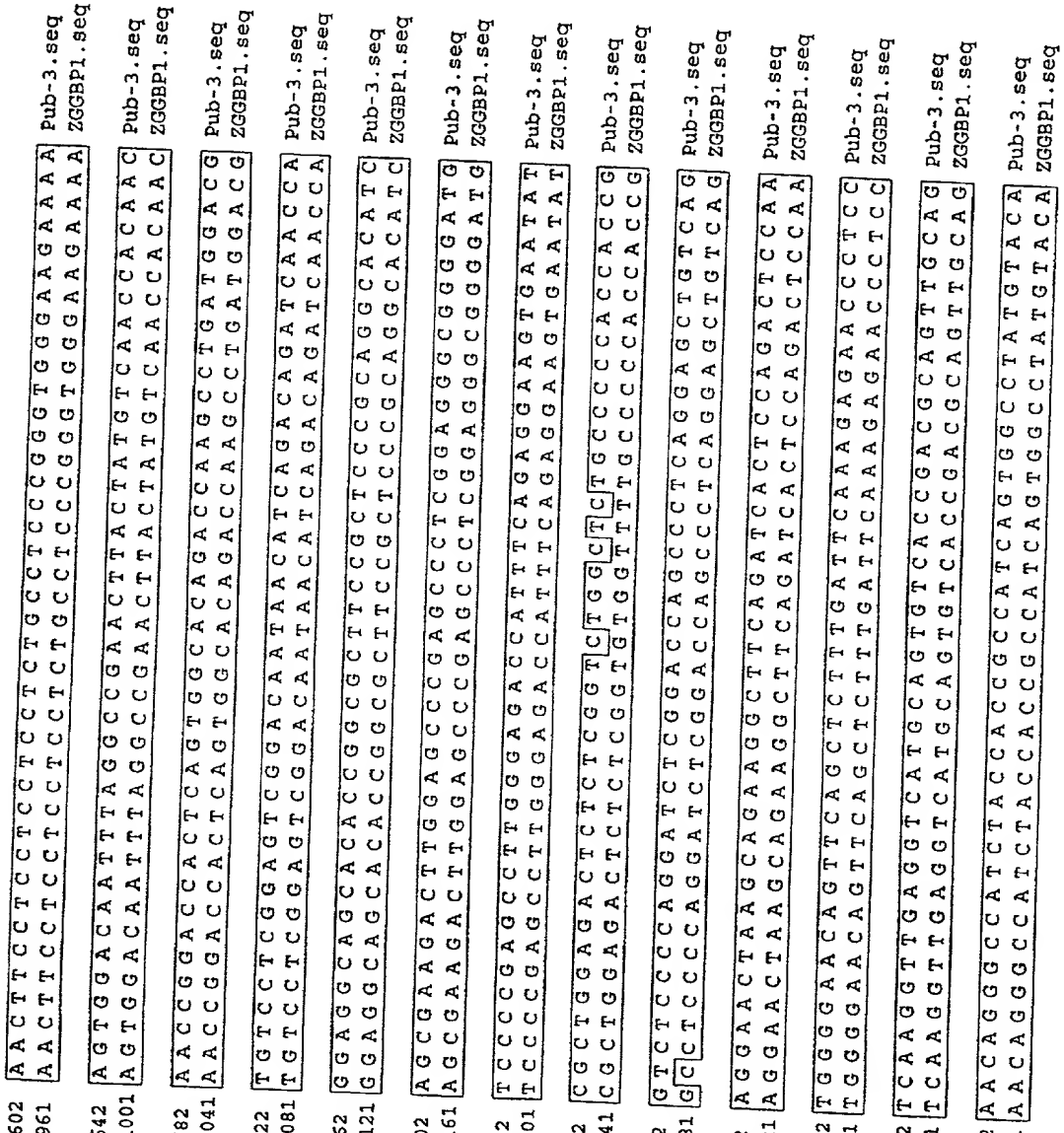


FIGURE 5 continued

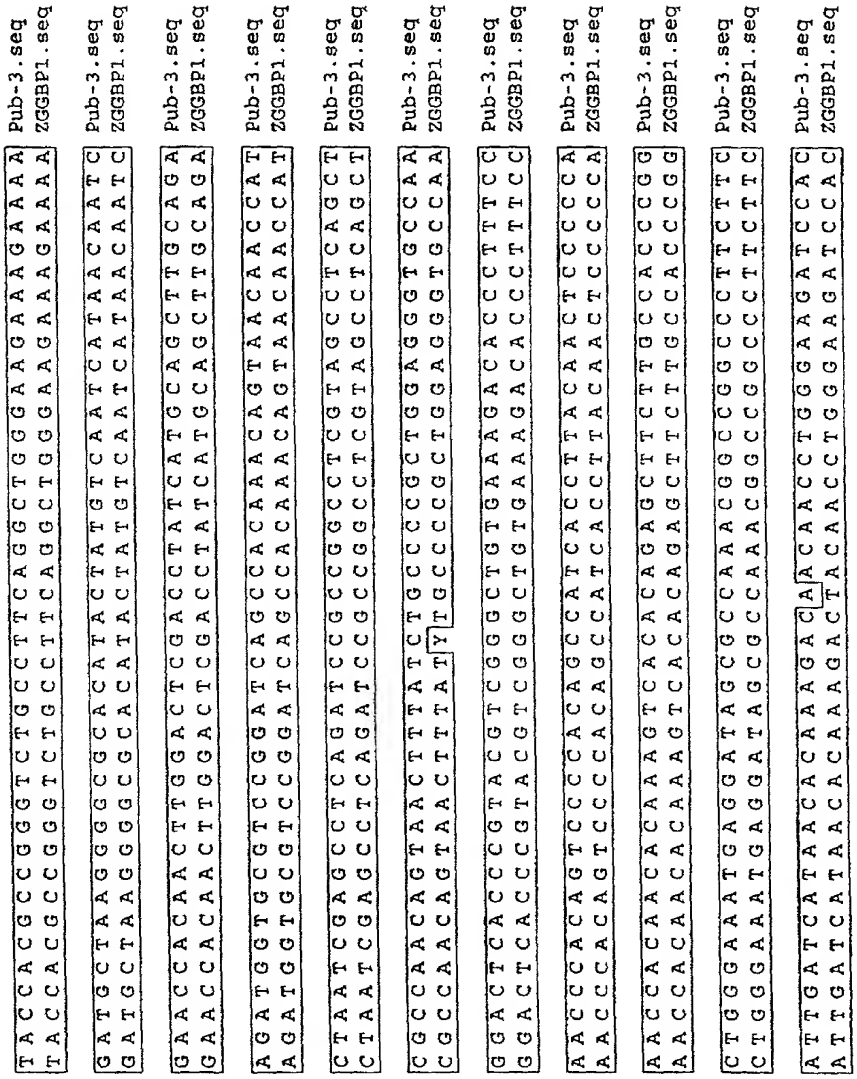


FIGURE 5 continued

```
1562 GTTTGAAATTTCCAGTACATATGCGGTCAAAGACATCTTT Pub-3.seq
1921 GTTTGAAATTTCCAGTACATATGCGGTCAAAGACATCTTT ZGGBP1.seq

1602 AAACCCCCAATGACCTTGGCCCCTTCCTCCTGGCTGGGAA Pub-3.seq
1961 AAACCCCAATGACCTTGGCCCCTTCCTCCTGGCTGGGAA ZGGBP1.seq

1642 GAAAGAATTCACTTGGATGGCCGAACGTTTTATATTGATC Pub-3.seq
2001 GAAAGAATTCACTTGGATGGCCGAACGTTTTATATTGATC ZGGBP1.seq

1682 ATAATAGCAAAATTACTCAGTGGGAAGACCCAAGACTGCA Pub-3.seq
2041 ATAATAGCAAAATTACTCAGTGGGAAGACCCAAGACTGCA ZGGBP1.seq

1722 GAACCCAGCTATTACTGGTCCGGCTGTCCCTTACTCCAGA Pub-3.seq
2081 GAACCCAGCTATTACTGGTCCGGCTGTCCCTTACTCCAGA ZGGBP1.seq

1762 GAATTTAAGCAGAAATATGACTACTTCAGGAAGAAATTAA Pub-3.seq
2121 GAATTTAAGCAGAAATATGACTACTTCAGGAAGAAATTAA ZGGBP1.seq

1802 AGAAACCTGCTGATATCCCCAATAGGTTTGAAATGAAACT Pub-3.seq
2161 AGAAACCTGCTGATATCCCCAATAGGTTTGAAATGAAACT ZGGBP1.seq

1842 TCACAGAAATAACATATTTGAAGAGTCCTATCGGAGAATT Pub-3.seq
2201 TCACAGAAATAACATATTTGAAGAGTCCTATCGGAGAATT ZGGBP1.seq

1882 ATGTCCGTGAAAAGACCAGATGTCCTAAAAGCTAGACTGT Pub-3.seq
2241 ATGTCCGTGAAAAGACCAGATGTCCTAAAAGCTAGACTGT ZGGBP1.seq

1922 GGATTGAGTTTGAATCAGAGAAAGGTCTTGACTATGGGGG Pub-3.seq
2281 GGATTGAGTTTGAATCAGAGAAAGGTCTTGACTATGGGGG ZGGBP1.seq

1962 TGTGGCCAGAGAATGGTTCTTCTTACTGTCCAAAGAGATG Pub-3.seq
2321 TGTGGCCAGAGAATGGTTCTTCTTACTGTCCAAAGAGATG ZGGBP1.seq

2002 TTCAACCCCTACTACGGCCTCTTTGAGTACTCTGCCACGG Pub-3.seq
2361 TTCAACCCCTACTACGGCCTCTTTGAGTACTCTGCCACGG ZGGBP1.seq

2042 ACAACTACACCCTTCAGATCAACCCTAATTCAGGCCTCTG Pub-3.seq
2401 ACAACTACACCCTTCAGATCAACCCTAATTCAGGCCTCTG ZGGBP1.seq
```

FIGURE 5 continued

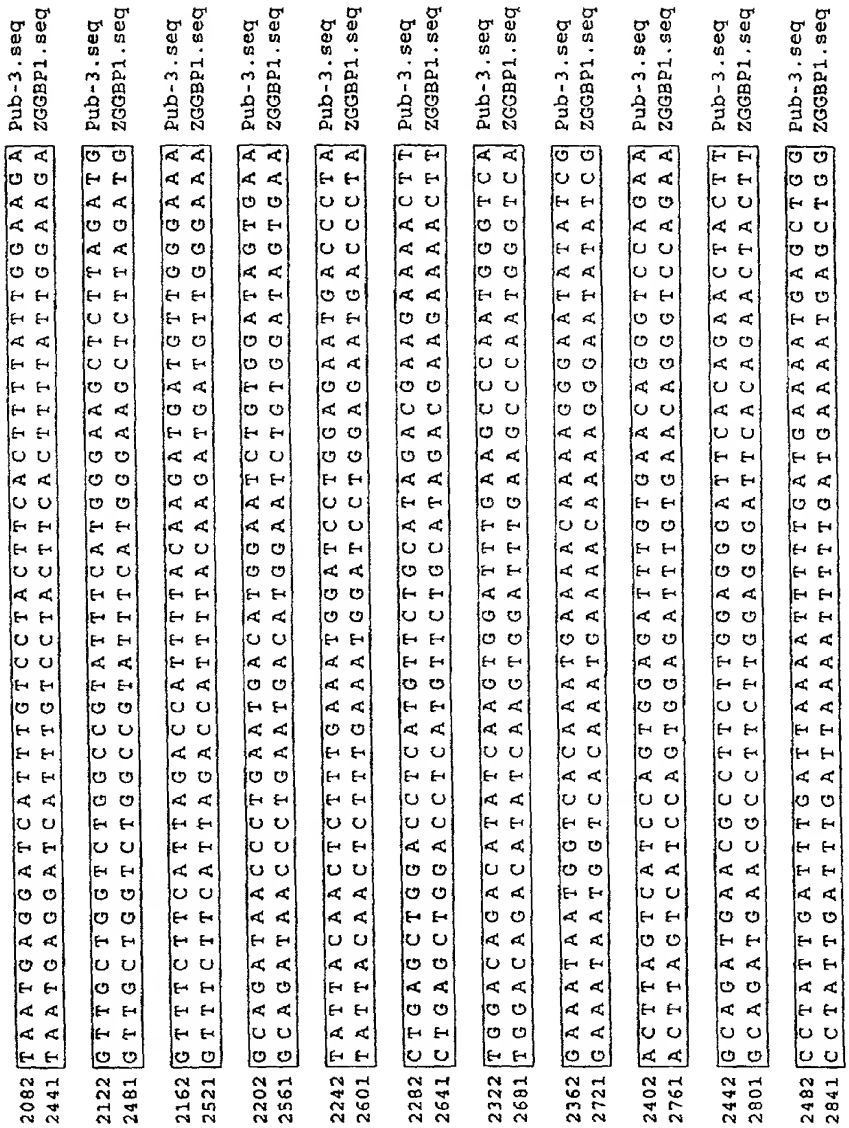


FIGURE 5 continued

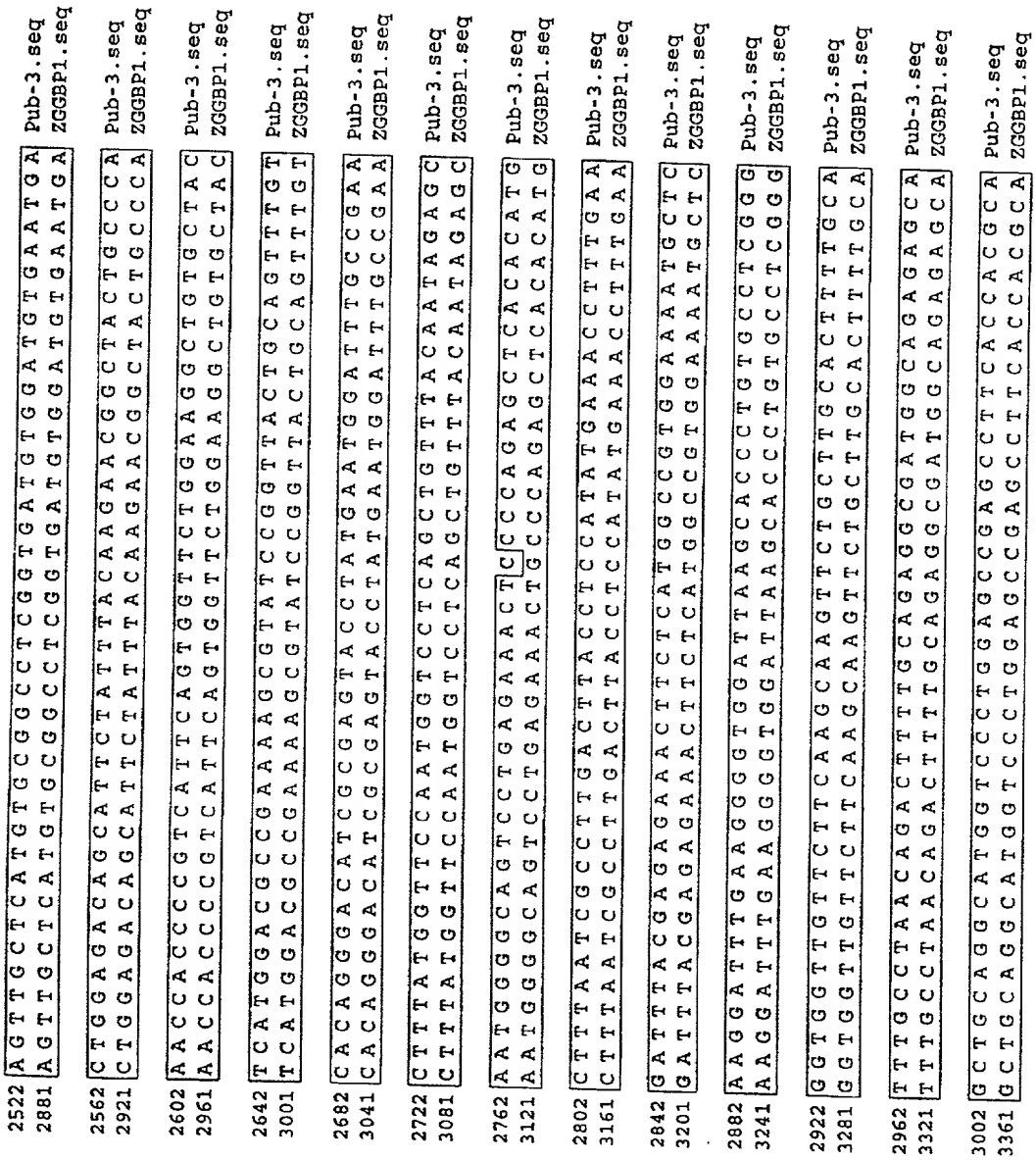


FIGURE 5 continued

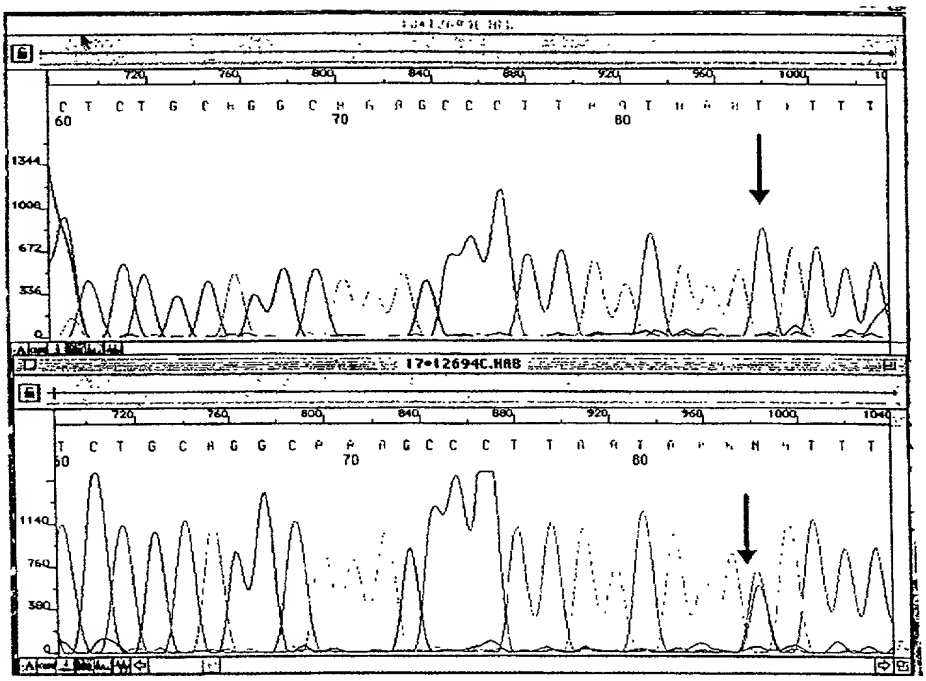
FIGURE 5 continued

```
3214 ----------------------------------------  Pub-3.seq
4360 TATCATGAACATTAAATGTGATGATGATTTCTTTTCCCTG  ZGGBP1.seq

3214 ----------------------------------------  Pub-3.seq
4400 CACACATCTTTCCGGTGCAATATCTATCAATTGTGAATCT  ZGGBP1.seq

3214 ----------------------------------------  Pub-3.seq
4440 GGCTGCTGGTGTATAAAAACCTGGATGTAAAGCTGAGCCT  ZGGBP1.seq

3214 ----------------------------------------  Pub-3.seq
4480 ACAGACCTGTCCTCACCAACTGTTTTGTGATTTCTACTCA  ZGGBP1.seq

3214 ----------------------------------------  Pub-3.seq
4520 ACTACAAAGATTTATTTAATGTACTCTTAATCTAACTGAG  ZGGBP1.seq

3214 ----------------------------------------  Pub-3.seq
4560 TTTTGTTACCAATGACCTGTTGCATGCTTCAATACCGTGT  ZGGBP1.seq

3214 ----------------------------------------  Pub-3.seq
4600 ACTGCCTGAGTTGTGCCTCTTGTGTGCTAGATTAAAAGTG  ZGGBP1.seq

3214 ----------------------------------------  Pub-3.seq
4640 AGACAGAGACTTGACTTGATCCTCTGAGCCTCAAGCTATT  ZGGBP1.seq

3214 ----------------------------------------  Pub-3.seq
4680 GAGCTGGTAGTGGCAGAGGACTGAGGGTACCTGCACAGTT  ZGGBP1.seq

3214 ----------------------------------------  Pub-3.seq
4720 TGATTCTTTTCCCACGTTGTAAGTCTCCATTGCAGAATTG  ZGGBP1.seq

3214 ----------------------------------------  Pub-3.seq
4760 TCGTGCGTTTGAGAAAACACCTGAGGCAGTGTGGGAGTTG  ZGGBP1.seq
```

FIGURE 5 continued

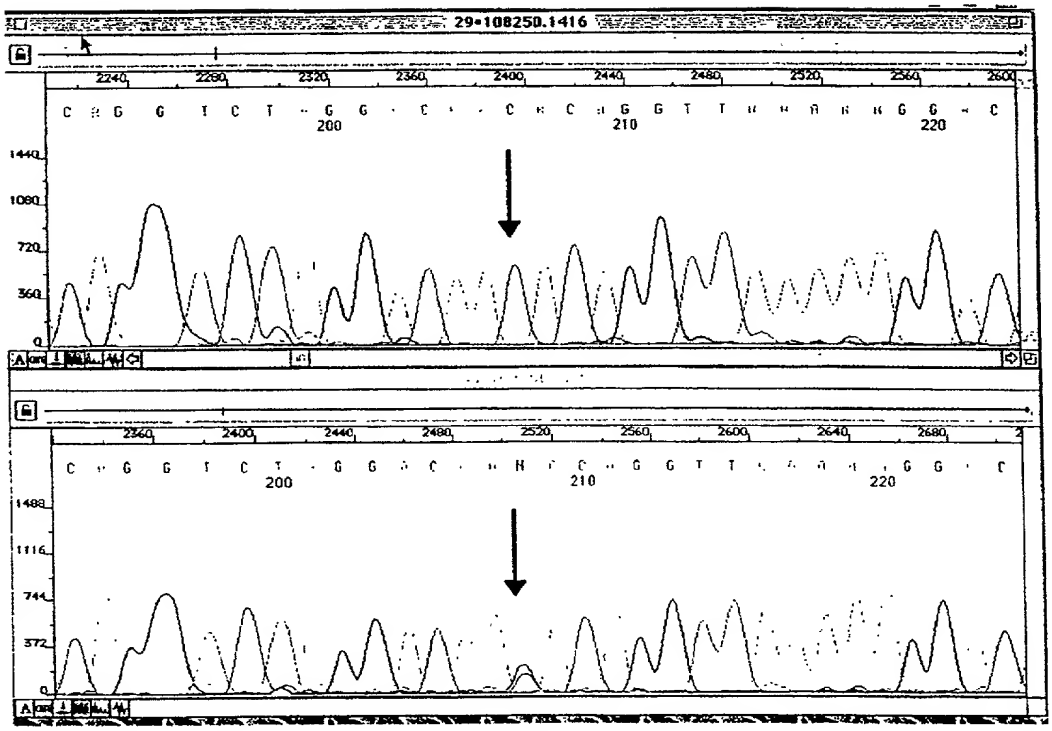
```
3214 ---------------------------------------- Pub-3.seq
4800 AACGACCCTGCTGTCCTTTTTAACCTGTGTTGTCCTAGAC ZGGBP1.seq

3214 ---------------------------------------- Pub-3.seq
4840 CTGTCGGGGCAGTCAGGGGACACTAGAGATTTGATCTCAT ZGGBP1.seq

3214 ------------------AAAAAA---------------- Pub-3.seq
4880 GCGAGTCATCAATAGGACAAAAAAGTTGTGGTTTGGGGAG ZGGBP1.seq

3220 ---------------------------------------- Pub-3.seq
4920 GTCTGTTTGTTACATAAAAAGGACCTTTCGGTGTAAGAAA ZGGBP1.seq

3220 ---------------------------------------- Pub-3.seq
4960 TTGCCGTTTTTACCCTGCCCTGGCTGGCATGTGAGAAGCC ZGGBP1.seq

3220 ---------------------------------------- Pub-3.seq
5000 ATGGAAGGTTGTGGTTGTAAATGAGTTGTCTAAAGGGGTG ZGGBP1.seq

3220 ------------------AAAAAAA                Pub-3.seq
5040 CAGAGGCCTGAGGTTTCTAAAAGAAGGTAGATTTCTACAG ZGGBP1.seq

3226                                          Pub-3.seq
5080 AGCTGAGTGTTGGTTCCTTTTTCTTATTGGTTGAAAATTA ZGGBP1.seq

3226                                          Pub-3.seq
5120 CCTGGTAGTGATCAGAAAACTTAGATGCTATGTAACTC   ZGGBP1.seq
```

Decoration 'Decoration #1': Box residues that match the Consensus exactly.



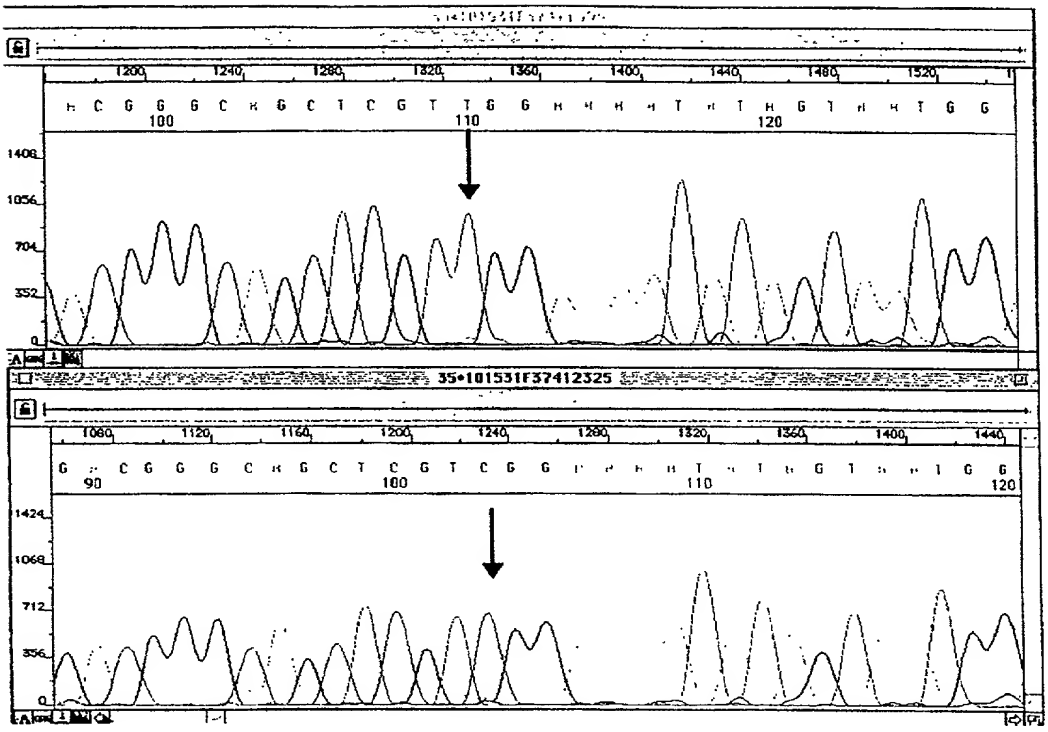
Wild Type (human foetal brain)	**T/T**
Variant Type (human adult brain)	**T/C**
Polymorphism Position	**3554**
RFLP	-

FIGURE 6



Wild Type (GM1416)	C/C
Variant (7225)	C/G
Position	4828

FIGURE 7



Primer sequences derived from BAC and used on lymphoblastoid cell lines from BPAD Patients.

Homozygous wild type (KK169) - T/T

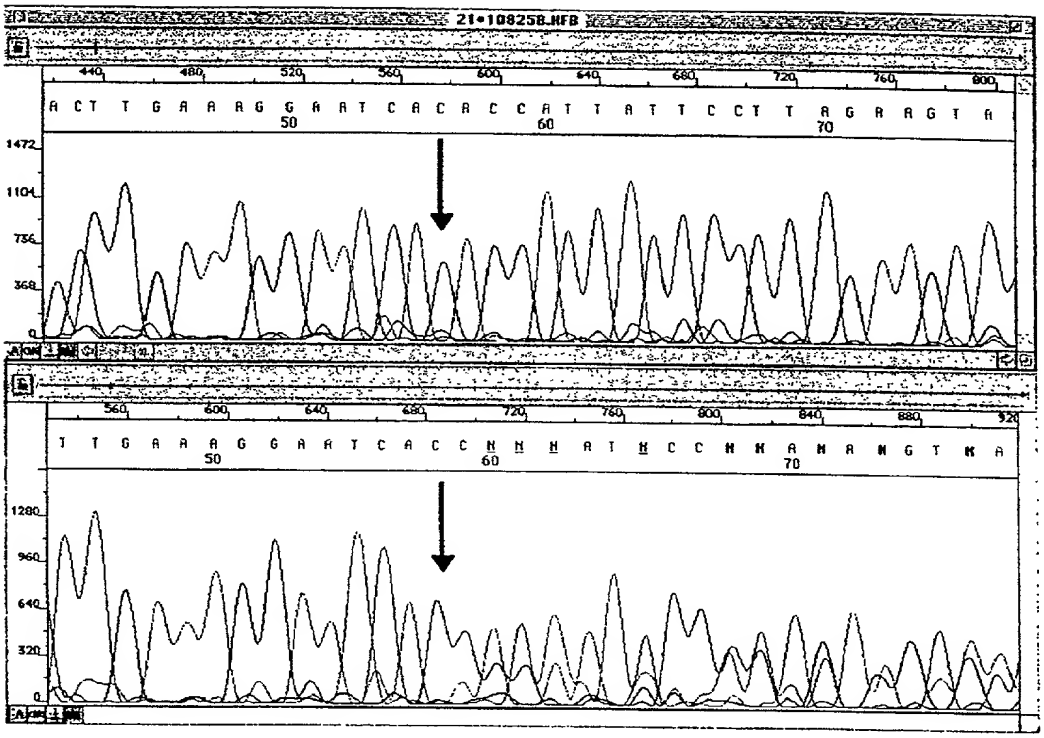
Homozygous variant (KK232) - C/C

FIGURE 8

Figure 9

TGCTGCAAGTGACAGGTTCCAAGAAGCCCGAGGGCTCAGAGCTGAATGATGAAGCGC
AGTCCCCAAAGTGCCTGGCCACCCCTCCCTCCCTGGATCACTGCTGCCTGGGCTTGA
TTGATTGATTGATTGATTGATTGATTGATTTTGAGAGAGATTCTCACTGTCACCCAG
GCTGGAGTACAGTGGTGCGATCTCGGCTCACTGCAGCCTCTGCCTCCCGGGTTCAAG
CAATTCTCCTGCCTCAGCCTCCCAAGTAGCTGGGACTACAGGCACGCGCCACCACAC
CCAGCTAATTTTGTATTTTTAGTAAAAGACGGGGTTTCACCATGTTGGGCCAGGATG
GTCTTGATCTCCTGACCTCATGATCCACCCGCCCCGGCTTCCAAAGTGCTGGGATAC
AGGCATGAACCCGACGCGCCCAGCATGGACATTTTTTTTTAATCCCCTGCCCTTTTC
TTGNGGCATAATTCATTGCAGGTCTCTTCTATACAGATCATGGAAAACACATTTTCT
TAACTGAGTTNTTATTATTTATACCCAGNCACCTCATGACANNTTTACCCTGTTACA
NACAAAATGGGCACCTGCCAAAANCAACTTTNATATAAGGATGCTCCAGGCCT

Tetranucleotide repeat underlined



Top electropherogram (human foetal brain) - wild type

Lower electropherogram (7225) - heterozygous variant

Arrow indicates the position of the C+C insertion - position 4032

FIGURE 10

RULE 63 (37 C.F.R. 1.63)
DECLARATION AND POWER OF ATTORNEY FOR PATENT APPLICATION IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

As a below named inventor, I hereby declare that my residence, post office address and citizenship are as stated below next to my name and I believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural names are listed below) of the subject matter which is claimed and for which a patent is sought on the INVENTION ENTITLED:

ZGGBP1, NOVEL PEPTIDES RELATED TO BIPOLAR AFFECTIVE DISORDERS TYPE 1, SEQUENCES AND USES THEREOF.

the specification of which

is attached hereto
was filed on as U.S. application serial No.
X was filed as PCT international application No. PCT/GB98/02259 on 28/07/1998
and (if applicable) was amended on

I hereby state that I have reviewed and understand the contents of the above identified specification, including the claims, as amended by any amendment referred to above. I acknowledge the duty to disclose all information which is known to me to be material to patentability as defined in 37 C.F.R. 1.56. I hereby claim foreign priority benefits under 35 U.S.C. 119/365 of any foreign application(s) for patent or inventor's certificate listed below and have also identified below any foreign application for patent or inventor's certificate filed by me or my assignee disclosing the subject matter claimed in this application and having a filing date (1) before that of the application on which priority is claimed, or (2) if no priority is claimed, before the filing date of this application:

PRIOR FOREIGN APPLICATION(S)

Number	Country	Day/MONTH/Year Filed	Date First Laid Open or published	Date Patented or Granted	Priority claimed Yes	Priority claimed No
9716162.4	United Kingdom	1 Aug 1997			X	

I hereby claim the benefit under 35 U.S.C. 120/365 of all United States applications listed below and PCT international applications listed above or below and, if this is a continuation-in-part (CIP)

application insofar as the subject matter disclosed and claimed in this application is in addition to that disclosed such in the prior applications. I acknowledge the duty to disclose all information known to me to be material to patentability as defined in 37 C.F.R. 1.56 which became available between the filing date of each such prior application and the national or PCT international filing date of this application:

PRIOR U.S. OR PCT APPLICATION(S)		Status
Application No. (Serial Code/Serial No.)	Day/MONTH/Year Filed	(patented, pending abandoned)

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

And I hereby appoint Cushman, Darby & Cushman, L.L.P. 1100 New York Avenue, N.W. Ninth Floor, East Tower, Washington, D.C. 20005-3918, telephone number 861-3000 (to whom all communications should be directed), and the below named persons (of the same address) individually and collectively my attorneys to prosecute this application and to transact all business in the Patent and Trademark Office connected therewith and with the resulting patent and I hereby authorise them to act and reply on instructions from and communicate directly with the person/assignee/attorney/firm/organisation who/which first sends/sent this case to them and by who/which I hereby declare that I have consented after full disclosure to be represented unless/until I instruct Cushman, Darby and Cushman in writing to the contrary.

Paul N Kokulis	16773	Kevin E Joyce	20508
Raymond F Lippitt	17519	Edward M Prince	22429
G Lloyd Knight	17698	James D. Berquist	34776
Carl G Love	18781	David W Brinkman	20817
Edgar H Martin	20534	George M Sirilla	18221
William K West Jr	22057	Timothy J. Klima	34852

W Warren Taltavull	25647	Donald J Bird	25323
Peter W Gowdey	25872	Lawrence Harbin	27644
Dale S Lazar	28872	Paul E White Jr	32011
Glenn J Perry	28458	Kendrew H Colton	30368
Chris Comuntzis	31097	Jeffrey A Simenauer	31933
Michelle N Lester	32331	John P. Moran	30906
Robert A Molan	29834	David A Jakopin	32955
G Paul Edgell	24238	Mark G Paulson	30793
Lynn E Eccleston	35861		

INVENTOR'S SIGNATURE Angela Veronica Flannery Date 16th December 1999
ANGELA VERONICA FLANNERY

Alderley Park,, Macclesfield, , Cheshire. SK10 4TG, United Kingdom. GBX
Residence and Post Office Address

British
Citizenship

INVENTOR'S SIGNATURE Maria Christina Martina Finnegan Date 20th December 1999
MARIA CHRISTINA MARTINA FINNEGAN

Alderley Park, Macclesfield, , Cheshire. SK10 4TG, United Kingdom. GBX
Residence and Post Office Address

British
Citizenship

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:
(A) NAME: Zeneca Limited
(B) STREET: 15 Stanhope Gate
(C) CITY: London
(D) STATE: England
(E) COUNTRY: United Kingdom
(F) POSTAL CODE (ZIP): W1Y 6LN
(G) TELEPHONE: 0171 304 5000
(H) TELEFAX: 0171 304 5151
(I) TELEX: 0171 304 2042

(ii) TITLE OF INVENTION: NOVEL COMPOUNDS

(iii) NUMBER OF SEQUENCES: 5

(iv) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

(vi) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: GB 9716162.4
(B) FILING DATE: 01-AUG-1997

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5154 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

CAAGCGCGCA ATTAACCCTC ACTAAAGGGA ACACCAACAC GTCGCCAGGA
CTGCGCCGTT 60

CGCTGCGCTC ATAGGCGGCG ATTTCATCAA GGGTGGCAAG GATCGCCTGG TCGACGGTCA 120

GGTCGTCCTC GACGCGGTTG CCCTCCTCGT CCTGTTCCAG GGTGAGTGGG CGATACCAGG 180

TGTCCACCGG GAAGGTACGG CCCGACACCT CGACAATCGG CGCATCGTCG AAGTGCTTGG 240

AAAAGCGCTC CAGGTCGATG GTGGCCGAGG TGATGATGAC TTTCAGGTCG GGGCGACGCG 300

GCAACAGGGT CTTGAGGTAG CCGAGCAGGA AGTCGATGTT CAGGCTGCGT TCGTGGGCTT 360

CGTCGACGAC AGGCTCGCGT TATGGCTCCG CTTTCTGCGG CTCTCCTACC CTGGCATGGT 420

GTGTGTGTGT GCCTGTGTGC TACGGAGAGT CCCGTATTCT CAGAGTAAAA GTTGTTCTGG 480

AATGATCTCG CCAAAAAGGA CATCTTTGGA GCCAGTGATC CGTATGTGAA ACTTTCATTG 540

TACGTAGCGG ATGAGAATAG AGAACTTGCT TTGGTCCAGA CAAAAACAAT TAAAAAGACA 600

CTGAACCCAA AATGGAATGA AGAATTTTAT TTCAGGGTAA ACCCATCTAA TCACAGACTC 660

CTATTTGAAG TATTTGACGA AAATAGACTG ACACGAGACG ACTTCCTGGG CCAGGTGGAC 720

GTGCCCCTTA GTCACCTTCC GACAGAAGAT CCAACCATGG AGCGACCCTA TACATTTAAG 780

GACTTTCTCC TCAGACCAAG AAGTCATAAG TCTCGAGTTA AGGGATTTTT GCGATTGAAA 840

ATGGCCTATA TGCCAAAAAA TGGAGGTCAA GATGAAGAAA ACAGTGACCA GAGGGATGAC 900

ATGGAGCATG GATGGGAAGT TGTTGACTCA AATGACTCGG CTTCTCAGCA CCAAGAGGAA 960

CTTCCTCCTC CTCCTCTGCC TCCCGGGTGG GAAGAAAAAG TGGACAATTT AGGCCGAACT 1020

TACTATGTCA ACCACAACAA CCGGACCACT CAGTGGCACA GACCAAGCCT
GATGGACGTG 1080

TCCTCGGAGT CGGACAATAA CATCAGACAG ATCAACCAGG AGGCAGCACA
CCGGCGCTTC 1140

CGCTCCCGCA GGCACATCAG CGAAGACTTG GAGCCCGAGC CCTCGGAGGG
CGGGGATGTC 1200

CCCGAGCCTT GGGAGACCAT TTCAGAGGAA GTGAATATCG CTGGAGACTC
TCTCGGTGTG 1260

GTTTTGCCCC CACCACCGGC CTCCCCAGGA TCTCGGACCA GCCCTCAGGA
GCTGTCAGAG 1320

GAACTAAGCA GAAGGCTTCA GATCACTCCA GACTCCAATG GGGAACAGTT
CAGCTCTTTG 1380

ATTCAAAGAG AACCCTCCTC AAGGTTGAGG TCATGCAGTG TCACCGACGC
AGTTGCAGAA 1440

CAGGGCCATC TACCACCGCC ATCAGTGGCC TATGTACATA CCACGCCGGG
TCTGCCTTCA 1500

GGCTGGGAAG AAAGAAAAGA TGCTAAGGGG CGCACATACT ATGTCAATCA
TAACAATCGA 1560

ACCACAACTT GGACTCGACC TATCATGCAG CTTGCAGAAG ATGGTGCGTC
CGGATCAGCC 1620

ACAAACAGTA ACAACCATCT AATCGAGCCT CAGATCCGCC GGCCTCGTAG
CCTCAGCTCG 1680

CCAACAGTAA CTTTATTGCC CCGCTGGAGG GTGCCAAGGA CTCACCCGTA
CGTCGGGCTG 1740

TGAAAGACAC CCTTTCCAAC CCACAGTCCC CACAGCCATC ACCTTACAAC
TCCCCCAAAC 1800

CACAACACAA AGTCACACAG AGCTTCTTGC CACCCGGCTG GGAAATGAGG
ATAGCGCCAA 1860

ACGGCCGGCC CTTCTTCATT GATCATAACA CAAAGACTAC AACCTGGGAA
GATCCACGTT 1920

TGAAATTTCC AGTACATATG CGGTCAAAGA CATCTTTAAA CCCCAATGAC
CTTGGCCCCC 1980

TTCCTCCTGG CTGGGAAGAA AGAATTCACT TGGATGGCCG AACGTTTTAT ATTGATCATA 2040

ATAGCAAAAT TACTCAGTGG GAAGACCCAA GACTGCAGAA CCCAGCTATT ACTGGTCCGG 2100

CTGTCCCTTA CTCCAGAGAA TTTAAGCAGA AATATGACTA CTTCAGGAAG AAATTAAAGA 2160

AACCTGCTGA TATCCCCAAT AGGTTTGAAA TGAAACTTCA CAGAAATAAC ATATTTGAAG 2220

AGTCCTATCG GAGAATTATG TCCGTGAAAA GACCAGATGT CCTAAAAGCT AGACTGTGGA 2280

TTGAGTTTGA ATCAGAGAAA GGTCTTGACT ATGGGGGTGT GGCCAGAGAA TGGTTCTTCT 2340

TACTGTCCAA AGAGATGTTC AACCCCTACT ACGGCCTCTT TGAGTACTCT GCCACGGACA 2400

ACTACACCCT TCAGATCAAC CCTAATTCAG GCCTCTGTAA TGAGGATCAT TTGTCCTACT 2460

TCACTTTTAT TGGAAGAGTT GCTGGTCTGG CCGTATTTCA TGGGAAGCTC TTAGATGGTT 2520

TCTTCATTAG ACCATTTTAC AAGATGATGT TGGGAAAGCA GATAACCCTG AATGACATGG 2580

AATCTGTGGA TAGTGAATAT TACAACTCTT TGAAATGGAT CCTGGAGAAT GACCCTACTG 2640

AGCTGGACCT CATGTTCTGC ATAGACGAAG AAAACTTTGG ACAGACATAT CAAGTGGATT 2700

TGAAGCCCAA TGGGTCAGAA ATAATGGTCA CAAATGAAAA CAAAAGGGAA TATATCGACT 2760

TAGTCATCCA GTGGAGATTT GTGAACAGGG TCCAGAAGCA GATGAACGCC TTCTTGGAGG 2820

GATTCACAGA ACTACTTCCT ATTGATTTGA TTAAAATTTT TGATGAAAAT GAGCTGGAGT 2880

TGCTCATGTG CGGCCTCGGT GATGTGGATG TGAATGACTG GAGACAGCAT TCTATTTACA 2940

AGAACGGCTA CTGCCCAAAC CACCCCGTCA TTCAGTGGTT CTGGAAGGCT GTGCTACTCA 3000

TGGACGCCGA AAAGCGTATC CGGTTACTGC AGTTTGTCAC AGGGACATCG CGAGTACCTA 3060

TGAATGGATT TGCCGAACTT TATGGTTCCA ATGGTCCTCA GCTGTTTACA ATAGAGCAAT 3120

GGGGCAGTCC TGAGAAACTG CCCAGAGCTC ACACATGCTT TAATCGCCTT GACTTACCTC 3180

CATATGAAAC CTTTGAAGAT TTACGAGAGA AACTTCTCAT GGCCGTGGAA AATGCTCAAG 3240

GATTTGAAGG GGTGGATTAA GCACCCTGTG CCTCGGGGGT GGTTGTTCTT CAAGCAAGTT 3300

CTGCTTGCAC TTTTGCATTT GCCTAACAGA CTTTTGCAGA GGCGATGGCA GAGAGCAGCT 3360

GCAGGCATGG TCCCTGGAGC CGAGCCTTCA CCACGCACTC GTCCAAGTTC GGGATGCGGG 3420

AACCTGGTCC CAGCTTGAGT TCCTGCCTTT CCCACCACAA ATTATCAACT GGTTGATGTG 3480

TACACTAATT ACATTTCAGG AGGACTTAAT GCTATTTATG TTGTCCTCTG CAGGCAAAGC 3540

CCTTAATAAA TATTTTACAT CCTTTCTAAT GACAATGAAT GGAATTAATC ACTCAACAGG 3600

TATAGTATTA CGACTCATGT TTACTTTTTA AAATGATTTA GACCGATTTT CAGATTTTAT 3660

TTCGTTATGA TTAAAGATGT CTCATGTACT TGGAAAAGTG AGCATTTTTT TTTTTTTTTG 3720

TATTTCACTT TCATACCAGG CTTAATGTCA ATGACATTTT TATTTTTGAA GTACTCTGAC 3780

ACCTCCACCC TCTACTTTAT TAGAATTGGA AGGCAAATTT TTGTCCAAAA ACCTACAGAC 3840

AAGTACTTTG AGAGAATTTC CAATATAATA TTAGACATAA TGATAATTTT TTCCATACTC 3900

AGAATGAAAA ACTGGATATT ACGTTTTTGT TTTGGGGTTT TTTTGTACAA ATTTAGCTAA 3960

TAGCTACAGG CTGAGAGAAT TGTAACATAG CATGACAAAT TTTGTGTTGA CTTGAAAGGA 4020

ATCACACCAT TATTCCTTAG AAGTAATTAC ATGTGTTCTA ACACATTTGA GACAGGGTTG 4080

GACTCCCATT TCTCATCCGA GAAATTACTT AACCCTTCCT GGGCGCTGTA CAGTCATCTT 4140

TTATTCTATT TCCTCTTTGC TGTTTGTAGT AGAGACATTT TGAATGAAAC TTGGCACTGC 4200

TTGATTCAAA ACTGTGGAAA CCAGATCTGT TTAGTCTCCT GTTTGTATGC GTTTGCTAAT 4260

GGTAGCTAAA TAACCAGTTT TTGTTGTAAA TGCACCAATT CTGAAGGCAC TTTATGTACT 4320

ACATGGAGGT CATATCTGGT TTTGTTTTTA TTTTTTTATC ATGAACATTA AATGTGATGA 4380

TGATTTCTTT TCCCTGCACA CATCTTTCCG GTGCAATATC TATCAATTGT GAATCTGGCT 4440

GCTGGTGTAT AAAAACCTGG ATGTAAAGCT GAGCCTACAG ACCTGTCCTC ACCAACTGTT 4500

TTGTGATTTC TACTCAACTA CAAAGATTTA TTTAATGTAC TCTTAATCTA ACTGAGTTTT 4560

GTTACCAATG ACCTGTTGCA TGCTTCAATA CCGTGTACTG CCTGAGTTGT GCCTCTTGTG 4620

TGCTAGATTA AAAGTGAGAC AGAGACTTGA CTTGATCCTC TGAGCCTCAA GCTATTGAGC 4680

TGGTAGTGGC AGAGGACTGA GGGTACCTGC ACAGTTTGAT TCTTTTCCCA CGTTGTAAGT 4740

CTCCATTGCA GAATTGTCGT GCGTTTGAGA AAACACCTGA GGCAGTGTGG GAGTTGAACG 4800

ACCCTGCTGT CCTTTTTAAC CTGTGTTGTC CTAGACCTGT CGGGGCAGTC AGGGGACACT 4860

AGAGATTTGA TCTCATGCGA GTCATCAATA GGACAAAAAA GTTGTGGTTT GGGGAGGTCT 4920

GTTTGTTACA TAAAAAGGAC CTTTCGGTGT AAGAAATTGC CGTTTTTACC CTGCCCTGGC 4980

TGGCATGTGA GAAGCCATGG AAGGTTGTGG TTGTAAATGA GTTGTCTAAA GGGGTGCAGA 5040

GGCCTGAGGT TTCTAAAAGA AGGTAGATTT CTACAGAGCT GAGTGTTGGT TCCTTTTTCT 5100

TATTGGTTGA AAATTACCTG GTAGTGATCA GAAAACTTAG ATGCTATGTA ACTC 5154

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 975 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

Met Phe Arg Leu Arg Ser Trp Ala Ser Ser Thr Thr Gly Ser Arg Tyr
1 5 10 15

Gly Ser Ala Phe Cys Gly Ser Pro Thr Leu Ala Trp Cys Val Cys Val
20 25 30

Pro Val Cys Tyr Gly Glu Ser Arg Ile Leu Arg Val Lys Val Val Ser
35 40 45

Gly Ile Asp Leu Ala Lys Lys Asp Ile Phe Gly Ala Ser Asp Pro Tyr
50 55 60

Val Lys Leu Ser Leu Tyr Val Ala Asp Glu Asn Arg Glu Leu Ala Leu
65 70 75 80

Val Gln Thr Lys Thr Ile Lys Lys Thr Leu Asn Pro Lys Trp Asn Glu
85 90 95

Glu Phe Tyr Phe Arg Val Asn Pro Ser Asn His Arg Leu Leu Phe Glu
100 105 110

Val Phe Asp Glu Asn Arg Leu Thr Arg Asp Asp Phe Leu Gly Gln Val
115 120 125

Asp Val Pro Leu Ser His Leu Pro Thr Glu Asp Pro Thr Met Glu Arg
130 135 140

Pro Tyr Thr Phe Lys Asp Phe Leu Leu Arg Pro Arg Ser His Lys Ser
145 150 155 160

Arg Val Lys Gly Phe Leu Arg Leu Lys Met Ala Tyr Met Pro Lys Asn
165 170 175

Gly Gly Gln Asp Glu Glu Asn Ser Asp Gln Arg Asp Asp Met Glu His
180 185 190

Gly Trp Glu Val Val Asp Ser Asn Asp Ser Ala Ser Gln His Gln Glu
195 200 205

Glu Leu Pro Pro Pro Pro Leu Pro Pro Gly Trp Glu Glu Lys Val Asp
210 215 220

Asn Leu Gly Arg Thr Tyr Tyr Val Asn His Asn Asn Arg Thr Thr Gln
225 230 235 240

Trp His Arg Pro Ser Leu Met Asp Val Ser Ser Glu Ser Asp Asn Asn
245 250 255

Ile Arg Gln Ile Asn Gln Glu Ala Ala His Arg Arg Phe Arg Ser Arg
260 265 270

Arg His Ile Ser Glu Asp Leu Glu Pro Glu Pro Ser Glu Gly Gly Asp
275 280 285

Val Pro Glu Pro Trp Glu Thr Ile Ser Glu Glu Val Asn Ile Ala Gly
290 295 300

Asp Ser Leu Gly Val Val Leu Pro Pro Pro Pro Ala Ser Pro Gly Ser
305 310 315 320

Arg Thr Ser Pro Gln Glu Leu Ser Glu Glu Leu Ser Arg Arg Leu Gln
325 330 335

Ile Thr Pro Asp Ser Asn Gly Glu Gln Phe Ser Ser Leu Ile Gln Arg
340 345 350

Glu Pro Ser Ser Arg Leu Arg Ser Cys Ser Val Thr Asp Ala Val Ala
355 360 365

Glu Gln Gly His Leu Pro Pro Pro Ser Val Ala Tyr Val His Thr Thr
370 375 380

Pro Gly Leu Pro Ser Gly Trp Glu Glu Arg Lys Asp Ala Lys Gly Arg
385 390 395 400

Thr Tyr Tyr Val Asn His Asn Asn Arg Thr Thr Thr Trp Thr Arg Pro
405 410 415

Ile Met Gln Leu Ala Glu Asp Gly Ala Ser Gly Ser Ala Thr Asn Ser
420 425 430

Asn Asn His Leu Ile Glu Pro Gln Ile Arg Arg Pro Arg Ser Leu Ser
435 440 445

Ser Pro Thr Val Thr Leu Xaa Ala Pro Leu Glu Gly Ala Lys Asp Ser
450 455 460

Pro Val Arg Arg Ala Val Lys Asp Thr Leu Ser Asn Pro Gln Ser Pro
465 470 475 480

Gln Pro Ser Pro Tyr Asn Ser Pro Lys Pro Gln His Lys Val Thr Gln
485 490 495

Ser Phe Leu Pro Pro Gly Trp Glu Met Arg Ile Ala Pro Asn Gly Arg
500 505 510

Pro Phe Phe Ile Asp His Asn Thr Lys Thr Thr Thr Trp Glu Asp Pro
515 520 525

Arg Leu Lys Phe Pro Val His Met Arg Ser Lys Thr Ser Leu Asn Pro
530 535 540

Asn Asp Leu Gly Pro Leu Pro Pro Gly Trp Glu Glu Arg Ile His Leu
545 550 555 560

Asp Gly Arg Thr Phe Tyr Ile Asp His Asn Ser Lys Ile Thr Gln Trp
565 570 575

Glu Asp Pro Arg Leu Gln Asn Pro Ala Ile Thr Gly Pro Ala Val Pro
580 585 590

Tyr Ser Arg Glu Phe Lys Gln Lys Tyr Asp Tyr Phe Arg Lys Lys Leu
595 600 605

Lys Lys Pro Ala Asp Ile Pro Asn Arg Phe Glu Met Lys Leu His Arg
610 615 620

Asn Asn Ile Phe Glu Glu Ser Tyr Arg Arg Ile Met Ser Val Lys Arg
625 630 635 640

Pro Asp Val Leu Lys Ala Arg Leu Trp Ile Glu Phe Glu Ser Glu Lys
645 650 655

Gly Leu Asp Tyr Gly Gly Val Ala Arg Glu Trp Phe Phe Leu Leu Ser
660 665 670

Lys Glu Met Phe Asn Pro Tyr Tyr Gly Leu Phe Glu Tyr Ser Ala Thr
675 680 685

Asp Asn Tyr Thr Leu Gln Ile Asn Pro Asn Ser Gly Leu Cys Asn Glu
690 695 700

Asp His Leu Ser Tyr Phe Thr Phe Ile Gly Arg Val Ala Gly Leu Ala
705 710 715 720

Val Phe His Gly Lys Leu Leu Asp Gly Phe Phe Ile Arg Pro Phe Tyr
725 730 735

Lys Met Met Leu Gly Lys Gln Ile Thr Leu Asn Asp Met Glu Ser Val
740 745 750

Asp Ser Glu Tyr Tyr Asn Ser Leu Lys Trp Ile Leu Glu Asn Asp Pro
755 760 765

Thr Glu Leu Asp Leu Met Phe Cys Ile Asp Glu Glu Asn Phe Gly Gln
770 775 780

Thr Tyr Gln Val Asp Leu Lys Pro Asn Gly Ser Glu Ile Met Val Thr
785 790 795 800

Asn Glu Asn Lys Arg Glu Tyr Ile Asp Leu Val Ile Gln Trp Arg Phe
805 810 815

Val Asn Arg Val Gln Lys Gln Met Asn Ala Phe Leu Glu Gly Phe Thr
820 825 830

Glu Leu Leu Pro Ile Asp Leu Ile Lys Ile Phe Asp Glu Asn Glu Leu
835 840 845

Glu Leu Leu Met Cys Gly Leu Gly Asp Val Asp Val Asn Asp Trp Arg
850 855 860

Gln His Ser Ile Tyr Lys Asn Gly Tyr Cys Pro Asn His Pro Val Ile
865 870 875 880

Gln Trp Phe Trp Lys Ala Val Leu Leu Met Asp Ala Glu Lys Arg Ile
885 890 895

Arg Leu Leu Gln Phe Val Thr Gly Thr Ser Arg Val Pro Met Asn Gly
900 905 910

Phe Ala Glu Leu Tyr Gly Ser Asn Gly Pro Gln Leu Phe Thr Ile Glu
915 920 925

Gln Trp Gly Ser Pro Glu Lys Leu Pro Arg Ala His Thr Cys Phe Asn
930 935 940

Arg Leu Asp Leu Pro Pro Tyr Glu Thr Phe Glu Asp Leu Arg Glu Lys
945 950 955 960

Leu Leu Met Ala Val Glu Asn Ala Gln Gly Phe Glu Gly Val Asp
965 970 975

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 854 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

ACAATGGGGG CGTGGCAGAG AATGGTTCTT CTTACTGTCC AAAGAGATGT TTAACCCCTA 60

CTATGGCCTC TTCGAGTACT CTGCCACGGA CAACTACACA CTTCAGATCA ATCCCAACTC 120

AGGCCTCTGT AATGAAGACC ATTTGTCCTA TTTCACCTTC ATTGGAAGAG TTGCTGGCCT 180

AGCGGTGTTT CATGGGAAAC TCTTAGATGG ATTCTTCATT CGACCATTCT ACAAGATGAT 240

GCTGGGGAAG CAGATAACGC TGAACGACAT GGAGTCCGTG GACAGCGAGT ACTACAACTC 300

TTTGAAGTGG ATCTTAGAAA ACGACCCCAC GGAACTTGAC CTCATGTTCT GCATAGACGA 360

GAGAACTTTG GGCAGACATA CCAAGTGGAT CTGAAGCCCA ACGGGTCAGA AATAATGGTA 420

ACCAATGAGA ACAAACGAGA ATACATTGAC TTAGTCATCC AGTGGAGATT TGTGAACAGG 480

GTCCAGAAGC AAATGAATGC CTTCTTGGAG GGATTTACAG AACTTCTTCC AATCGACTTG 540

ATTAAAATTT TTGATGAAAA TGAGCTGGAG TTGCTGATGT GCGGCCTTGG TGATGTCGAC 600

GTGAACGACT GGAGACAGCA CTCTATTTAC AAGAACGGCT ACTGCCCCAA CCACCCTGTC 660

ATCCAGTGGT TCTGGAAGGC CGTGCTCCTG ATGGATGCTG AGAAGCGCAT CCGGTTACTA 720

CAGTTTGTCA CAGGCACCTC CAGAGTACCC ATGAATGGAT TTGCCGAACT CTATGGTTCC 780

AATGGTCCTC AGCTGTTTAC AATAGAGCAA TGGGGCAGTC CGAAAAACTA CCAGAGCTCT 840

ACATGCTTAA TCGC 854

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 604 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

His Ala Cys Ser Asn Ala Ala Ser Arg Ala Ala Ala Arg Val Ala Ala
1 5 10 15

Arg Cys Thr Ala Arg Ser Arg Ser Gly Arg Arg Ser Ser Ser Val Ser
20 25 30

Arg Ser Ser Ser Arg Gly Ala Ser Ser Ser Met Ser Ser Asp Met Ala
35 40 45

Ala Asp Ser Ala Val Ser Asp Val Trp Cys Asp Lys Thr Asp Gly Gly
50 55 60

Gly Ser Gly Ser Asp Val Thr Asp Thr Cys Cys Gly Cys Trp Asn Asn
65 70 75 80

Ser His Val Thr Ala Asp Tyr His Asn Asp Asp Thr Arg Val Val Arg
85 90 95

Val Lys Val Ala Gly Gly Ala Lys Lys Asp Gly Ala Ser Asp Tyr Val
100 105 110

Arg Val Thr Tyr Asp Met Ser Gly Thr Ser Val Thr Lys Thr Lys Lys
115 120 125

Ser Asn Lys Trp Asn Arg Val Arg His Arg Val Asp Asn Arg Thr Arg
130 135 140

Asp Asp Gly Val Asp Val Tyr Thr Asn Arg Met Arg Tyr Thr Lys Asp
145 150 155 160

Val His Arg Ser His Lys Ser Arg Val Lys Gly Tyr Arg Lys Met Thr
165 170 175

Tyr Lys Asn Gly Ser Asp Asn Ala Asp Ala Gly Trp Val Val Asp Asp
180 185 190

Ala Ala Thr His His Ser Gly Trp Arg Asp Val Gly Arg Thr Tyr Tyr
195 200 205

Val Asn His Ser Arg Arg Thr Trp Lys Arg Ser Asp Asp Asp Thr Asp
210 215 220

Asp Asn Asp Asp Met Ala Arg Ala Thr Thr Arg Arg Ser Asp Val Asp
225 230 235 240

Gly Asp Asn Arg Ser Asn Trp Val Arg Asp Asn Thr Tyr Ser Gly Ala
245 250 255

Val Ser Ser Gly His Asp Val Thr His Ala Asn Thr Arg Ala Val Cys
260 265 270

Gly Asn Ala Thr Ser Val Thr Ser Ser Asn His Ser Ser Arg Gly Gly
275 280 285

Ser Thr Cys Thr Val Thr Ser Ser Gly Gly Trp Lys Asp Asp Arg Gly
290 295 300

Arg Ser Tyr Tyr Val Asp His Asn Ser Lys Thr Thr Thr Trp Ser Lys
305 310 315 320

Thr Met Asp Asp Arg Ser Lys Ala His Arg Gly Lys Thr Asp Ser Asn
325 330 335

Asp Gly Gly Trp Arg Thr His Thr Asp Gly Arg Val Asn His Asn Lys
340 345 350

Lys Thr Trp Asp Arg Asn Val Ala Thr Gly Ala Val Tyr Ser Arg Asp
355 360 365

Tyr Lys Arg Lys Tyr Arg Arg Lys Lys Lys Thr Asp Asn Lys Met Lys
370 375 380

Arg Arg Ala Asn Asp Ser Tyr Arg Arg Met Gly Val Lys Arg Ala Asp
385 390 395 400

Lys Ala Arg Trp Asp Gly Lys Gly Asp Tyr Gly Gly Val Ala Arg Trp
405 410 415

Ser Lys Met Asn Tyr Tyr Gly Tyr Ser Ala Thr Asp Asn Tyr Thr Asn
420 425 430

Asn Ser Gly Cys Asn Asp His Ser Tyr Lys Gly Arg Val Ala Gly Met
435 440 445

Ala Val Tyr His Gly Lys Asp Gly Arg Tyr Lys Met Met Lys Thr His
450 455 460

Asp Met Ser Val Asp Ser Tyr Tyr Ser Ser Arg Trp Asn Asp Thr Asp
465 470 475 480

Arg Asp Gly Thr His His Lys Thr Gly Gly Ser Val Val Thr Asn Lys
485 490 495

Asn Lys Lys Tyr Tyr Val Trp Arg Val Asn Arg Lys Met Ala Ala Lys
500 505 510

Gly Asp Lys Asp Asn Met Cys Gly Gly Asp Val Asp Val Asn Asp Trp
515 520 525

Arg His Thr Lys Tyr Lys Asn Gly Tyr Ser Met Asn His Val His Trp
530 535 540

Trp Lys Ala Val Trp Met Met Asp Ser Lys Arg Arg Val Thr Gly Thr
545 550 555 560

Xaa Ser Arg Val Met Asn Gly Ala Tyr Gly Ser Asn Gly Ser Thr Val
565 570 575

Trp Gly Thr Asp Lys Arg Ala His Thr Cys Asn Arg Asp Tyr Ser Asp
580 585 590

Trp Asp Lys Met Ala Asn Thr Gly Asp Gly Val Asp
595 600

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 615 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

TGCTGCAAGT GACAGGTTCC AAGAAGCCCG AGGGCTCAGA GCTGAATGAT GAAGCGCAGT 60

CCCCAAAGTG CCTGGCCACC CCTCCCTCCC TGGATCACTG CTGCCTGGGC TTGATTGATT 120

GATTGATTGA TTGATTGATT GATTTTGAGA GAGATTCTCA CTGTCACCCA GGCTGGAGTA 180

CAGTGGTGCG ATCTCGGCTC ACTGCAGCCT CTGCCTCCCG GGTTCAAGCA ATTCTCCTGC 240

CTCAGCCTCC CAAGTAGCTG GGACTACAGG CACGCGCCAC CACACCCAGC TAATTTTGTA 300

TTTTTAGTAA AAGACGGGGT TTCACCATGT TGGGCCAGGA TGGTCTTGAT CTCCTGACCT 360

CATGATCCAC CCGCCCCGGC TTCCAAAGTG CTGGGATACA GGCATGAACC CGACGCGCCC 420

AGCATGGACA TTTTTTTTTA ATCCCCTGCC CTTTTCTTGG GCATAATTCA TTGCAGGTCT 480

CTTCTATACA GATCATGGAA AACACATTTT CTTAACTGAG TTTTATTATT TATACCCAGC 540

ACCTCATGAC ATTTACCCTG TTACAACAAA ATGGGCACCT GCCAAAACAA CTTTATATAA 600

GGATGCTCCA GGCCT 615